# THE EVOLUTION OF HAEMOGLOBIN GENE LOCI IN AMNIOTES

**Vidushi S. Patel**

Bachelor of Biotechnology (Hons I)

A thesis submitted for the degree of Doctor of Philosophy of

The Australian National University

March, 2011

# DECLARATION

Except where specific reference is made to other sources, the work presented in this thesis is the work of the author. It has not been submitted, in whole or in part for any other degree.

……

Vidushi S. Patel

# ACKNOWLEDGEMENTS

My heartful thanks firstly goes to Prof. Jenny Graves for taking me on as her student in her laboratory and providing me with ample support, inspiration and guidance throughout my PhD programme. She is the most ideal guru that a student could wish for. I am very honoured to have her as my supervisor and greatly appreciate her patience and trust in me to complete my PhD. Thank you Jenny for everything, including knowledge that you have given me. I will cherish these moments forever.

Another big thanks goes to my other-half supervisor, Dr Janine Deakin, without whom my PhD would not have been possible to complete. Her enthusiastic and encouraging supervision and comments were of great help and guidance to me. Janine was always approachable and understanding, which made my stay in a lab more joyous and memorable one. Oh, not to forget the wonderful pies she makes. I am looking forward to her pie night that is still pending once I finish my PhD.

I would like to thank all my collaborators and co-authors for their time and involvement in the projects. Their intellectual input and discussions have always been fruitful and inspiring. In particular, thank you to Dr Steven Cooper from The University of Adelaide for your assistance and expertise in phylogenetic studies. Thank you also to Assistant Prof. Tariq Ezaz from the University of Canberra for providing me with dragon lizard metaphases and helping me with FISH experiments. My sincere thanks to my international collaborator, Prof Ross Hardison, from Pennsylvania State University for your warm hospitality in your laboratory and assistance in the regulatory project. Thank you all for taking out your valuable time to teach me new techniques. In addition, thank you to all who have kindly donated animal tissues or DNA that has made this project possible: Ke-Jun Wei (for marsupial), Lauren Marrotte (tammar wallaby), Frank Grützner and Tim Hore (platypus), Travis Glenn and Jan Danes (anole lizard) and Lee Miles (for transporting anole gDNA).

My sincere thanks also to my lab members, who have enjoyed all occasions with me and gave their shoulders for me to cry on when my experiments were unsuccessful: Shafagh Al-Nadaf, Hannah Bender, Julie Chaumeil, Margaret Delbridge, Melanie Edwards, Nerida Harley, Barbara Harriss, Tim Hore, Kristen Jordan, Ali Livernois,

Daniel McMillan, Amir Mohammadi, Liz Murchison, Veronica Murtagh, Denis O'Meally, Hardip Patel, Claudia Rodrigues Delgado, Carly Smith, Paul Waters, and Ke-Jun Wei. Thank you all for your friendship and making the lab environment more of a playground rather than a classroom. My stay in the lab was really an enjoyable one.

Lastly but not the least, I would like to thank my family members who have provided me with invaluable time, support, patience and love throughout my student life. I am incredibly grateful to my parents, Vinod bhai and Jagruti ben Patel, for always being there for me and providing me with everything I needed. Although they live in Fiji, they have never missed a single birthday party of mine in Australia or any other auspicious occasions, including the birth of my first baby (and special thanks to my mum for making my thesis writing possible by taking out her precious time out to look after him). I have achieved so many things with your love and blessings and I thank you both. Thank you to my sibling and his family- Sudhish, Dhara and Aarushi Patel, and to all other family members:- Sanjit, Ripal, Bhumi, Shruti, Ashwani, Shaunak, Shiwani and Hinal especially for inspiring me to complete my thesis.

Finally, my heartful thanks goes to my supporting and understanding husband, Shaishav Patel, who has tolerated my tough days, late nights and stressful deadlines. With your love and patience, I am now a happy mum of Master Tanish Patel (now 10 months old) who means the whole world to me. You have always been there for me and I am very lucky to have you and Tanish in my life.

# ABSTRACT

The genes in alpha (α)- and beta (β)- globin clusters constitute a functional haemoglobin molecule, crucial for oxygen transportation. In most fish and amphibians, α- and β-globin genes are located together, whereas in amniotes (birds and mammals), there are two distinct clusters. Several complex models have been proposed to explain the evolution of these gene clusters. However, there was a lack of data for key positions in amniote phylogeny to discern which one was most parsimonious. Therefore, the main aims of this project were to characterise α- and β-globin clusters and their regulatory regions in a monotreme *Ornithorhynchus anatinus* (Australian duck-billed platypus) and two reptilian species *Pogona vitticeps* (Australian bearded dragon) and *Anolis carolinensis* (green anole lizard), to gain insight into globin loci evolution. This thesis is presented as a collection of research papers covering each topic, and a review and discussion that summarises my research.

The first paper (Chapter 2) reports a comprehensive study on the characterisation, expression and evolution of α- and β-globin gene clusters in the platypus, using a combination of molecular and bioinformatics approaches. The most important findings from this work leading to the development of a new and simple model for globin gene evolution concerned the discovery of a β-like globin gene within the α-globin cluster and genomic context analysis of α- and β-globin clusters across vertebrates. I showed that the amniote α-globin cluster is in fact the same as the α-β cluster found in fish and amphibians, and both clusters share common flanking genes (*C16orf35* and *LUC7L*). I proposed a transposition model in which a copy of β-globin gene was inserted into a cluster of olfactory receptors (flanked by *RRM1*, *CCKBR* and *ILK*) in the ancestor of amniotes, thus originating the amniote β-globin cluster. To elaborate this model further, my second paper (Chapter 3) reviews some events that could have led to this transposition, and their effects on the current fate of regulation.

Information on the organization of globin genes in reptiles was required to test this transpositional model. I looked into the globin gene organization in the green anole using a bioinformatics approach and in the bearded dragon using a molecular approach. The results are reported in Chapter 4 and my third paper, which describe how

fragmentary data from the green anole genome sequence assembly and mapping data from bearded dragon provided further evidence to support my proposed model for the evolution of the β-globin gene cluster in amniotes.

I also studied the evolution of regulatory regions of the platypus α- and β-globin clusters to address the question whether the translocation of the β-globin locus resulted in a transposition of its regulatory region, or whether a new regulatory region evolved as a result of this translocation (reported in the fourth paper, Chapter 5). By using some novel techniques, I showed that the platypus α-globin has a major regulatory element that is conserved with other jawed vertebrates, whereas the regulatory regions of their β-globin cluster do not show any conservation at the sequence level to those of birds and therian mammals. This suggested that the regulatory regions of amniote β-globin genes evolved either more rapidly (more substitutions) or more extensively (e.g. more rearrangements) from a common ancestral regulatory region. Alternatively, these regulatory regions may have independent origins in different amniote lineages.

In my final chapter, I discuss the overall implications of my findings on this area of research. I highlight the special value of studying non-model species mammals and reptiles, by which researchers are able to gain novel information about globin evolution and regulation.

# PUBLICATIONS

**Patel, V.S.**, Dorman, C., King, D.C., et al. (in prep) *Cis*-regulatory regions of the platypus α- and β-globin loci and their evolution. *Molecular Biology and Evolution*

**Patel, V.S.,** Ezaz, T., Deakin, J.E., Graves, J.A.M. (2010) Globin gene structure in a reptile supports the transpositional model for amniote alpha- and beta- globin gene evolution. *Chromosome Res* **18:** 897-907

**Patel, V.S.,** and Deakin, J.E. (2010) The evolutionary history of globin genes: insights from marsupials and monotremes. Book review *Marsupial Genetics and Genomics*. Ed: Deakin, Waters and Graves: 415-433

**Patel, V.S.,** Cooper, S.J.B., Deakin, J.E., et al. (2008) Platypus globin genes and flanking loci suggest a new insertional model for β-globin evolution in birds and mammals. *BMC Biology* **6:34**

Deakin, J.E., Koina, E., Waters, P.D., Doherty, R., **Patel, V.S.,** et al. (2008) Physical map of two tammar wallaby chromosomes: a strategy for mapping in non-model mammals. *Chromosome Res* **16(8)**: 1159-1175

# LIST OF ABBREVIATIONS

There has been inconsistency in referring to globin gene names in various species and literature. Previous studies referred to α- and β-globin genes based on their Greek names but recent studies refer to them according to standard nomenclature, as specified by Aquileta et al. (2006b) shown below. My first publication (Chapter 2) uses Greek names, whilst the other chapters adopt standard nomenclature to avoid confusion. **Note:** I use generic 'α-globin' or 'β-globin' to refer to the entire globin clusters that contain many 'α-like' and β-like' globin genes but I use the standard nomenclature indicated below in italics when I refer to a specific gene contained in these clusters. I also use Greek names for fish and frog globin genes because they are les well characterised then amniote globin genes. A gene name followed by '–ps' refers to a pseudogene.

| **Greek names** | **Standard Nomenclature** |
|---|---|
| π | *HBP* |
| ζ | *HBZ-T1* (subunit 1 of HBZ) |
| ζ′ | *HBZ-T2* (subunit 2 of HBZ) |
| $\alpha^D$ or μ | *HBK* |
| $\alpha^A$ | *HBA* |
| $\alpha^1$ | *HBA-T1* (subunit 1 of HBA) |
| $\alpha^2$ | *HBA-T2* (subunit 2 of HBA) |
| $\alpha^3$ | *HBA-T3* (subunit 3 of HBA) |
| θ | *HBQ* |
| ω | *HBW* |
| ε | *HBE* |
| γ | *HBG* |
| δ | *HBD* |
| β | *HBB* |
| $\beta^H$ | *HBB-T1* (subunit 1 of HBB) |
| $\beta^A$ | *HBB-T2* (subunit 2 of HBB) |
| βI | *HBB1* (refers to βI chain) |
| βII | *HBB2* (refers to βII chain) |
| η | *HBH* |

# TABLE OF CONTENTS

# CHAPTER 1: INTRODUCTION

## Comparative Genomics

The new era for DNA technology began with the sequencing of our own species, *Homo sapiens*, in order to decode the genetic make-up and to understand human development, physiology, medicine and evolution (2001). However, the human genome sequence, while an extremely valuable resource, was difficult to interpret on its own and the need to sequence genomes of other vertebrate species for comparisons was identified as a crucial step for annotation of the human genome and to facilitate studies into genome evolution (Green, 2007).

Since then, more than 50 vertebrate species have been sequenced, with some receiving high coverage (>6 times) and others receiving low coverage (less than 2 times). As of release 60 (8 November 2010), Ensembl supports 56 species, including five preliminarily (Flicek et al., 2010) and the numbers are increasing rapidly. This enables researchers to perform multi-species sequence alignment, which can reveal evolutionarily conserved (presumably functional) sequences shared by closely- and distantly- related species. Genome sequencing technology has, therefore, created the foundation for comparative genomics. Armed by increasingly powerful bioinformatics tools, researchers are now beginning to understand how the genome is functionally interpreted, and how it evolved over time.

Globin genes are classic examples of the duplication and diversification of genes, and provide an ideal model for the use of comparative genomics to understand function and evolution. For example, comparative analysis of globin genes in multiple species can help identify, as well as conserved genes, regulatory regions that control the timing and expression of these globin genes during development. It can also allow us to reconstruct the sequence of an ancestral globin gene in an ancestral vertebrate, and to trace changes that occurred in the evolution of specialised forms and functions of globins. This would then help us explain how duplicate genes acquired different functions, and provide information about mechanisms affecting their regulation. This information, for example in haemoglobin (Hb) genes, can then facilitate potential medical and clinical

applications to test and treat diseases associated with globin abnormalities. Thus, it is of great interest to compare the structure and sequence of globin genes in all vertebrates.

In this chapter, I present an up-to-date summary on vertebrate globins, with an emphasis on Hb genes. Details of human Hb genes, their synthesis, regulation and diseases associated with globin abnormalities, are also presented in order to understand the importance of such genes and proteins to us and to medical research. Next, the structure, organization and regulatory regions of Hb genes are compared across jawed vertebrates previously studied to chart the evolutionary history of Hb genes and to identify gaps in the evolutionary record. Finally, I list my aims, and present an outline of the manuscripts and published articles that make up the body of this thesis.

## Globin diversity

Globins are a family of small haem-proteins that bind oxygen and other gaseous ligands between the iron-ion of the porphyrin ring and usually a histidine of the polypeptide chain (Dickerson and Geis, 1983). There are many large and diverse families of globins. Two are particularly well-studied models for protein functions and gene family evolution (Dickerson and Geis, 1983, Goodman et al., 1987); myoglobin (Mb), which either transports or stores oxygen in the myomers of the striated and cardiac muscle, also detoxifies nitric oxide (Wittenberg and Wittenberg, 2003), and Hb, which binds and transports oxygen between tissues. More recently, other globin gene families have been discovered, including cytoglobin (Cygb), neuroglobin (Ngb), globin E (GbE), globin X (GbX), globin Y (GbY). These add further interest to globin studies, although little is known yet of their physiological functions (Burmester et al., 2000, Kawada et al., 2001, Burmester et al., 2002, Trent and Hargrove, 2002, Kugelstadt et al., 2004, Roesner et al., 2005, Fuchs et al., 2006).

Phylogenetic studies group all globins into two distinct clades. The globins in one clade, containing Ngb and GbX, are ancient and are more closely related to annelid intracellular globins than to other vertebrate globins. The other clade contains vertebrate-specific globins such as Mb, Cygb, GbE, GbY and Hbs of jawed vertebrates (gnathostomes) and jawless fish (cyclostomes) (Burmester et al., 2000, Burmester et al., 2004, Roesner et al., 2005, Fuchs et al., 2006).

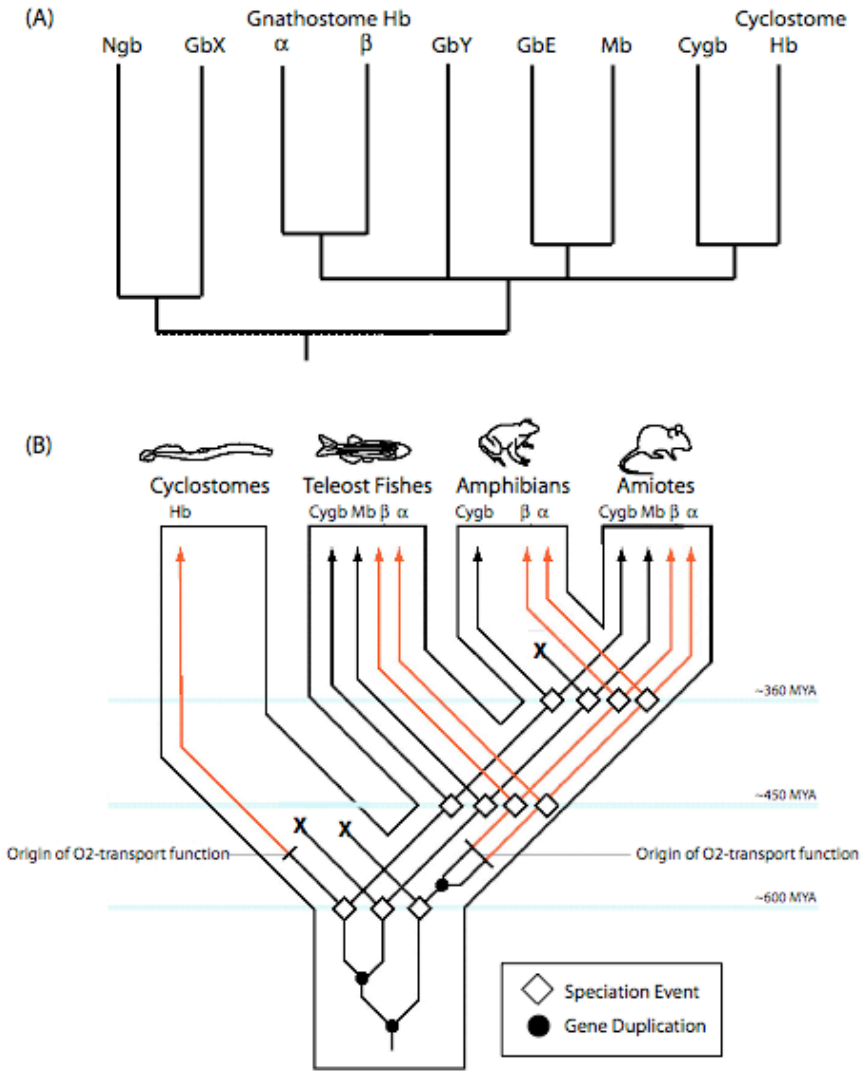



**Figure 1: Evolution of vertebrate-specific globins**

(A) Phylogenetic relationship between neutroglobin (Ngb), Cytoglobin (Cygb), Myoglobin (Mb), Globin X (GbX), Globin Y (GbY), Globin E (GbE), and Haemoglobin (Hb) of gnathostomes and cyclostomes grouped them into two distinct clades. Adapted from Hoffmann et al. (2010a); (B) Detailed model showing the evolution of Hb, Mb and Cygb in vertebrates. According to this model the ancestral single-copy globin gene at the stem of the vertebrate lineage duplicated to produce two descendent gene lineages; one gave rise to Hb (α and β) in all gnathostomes (*left branch*) but was secondarily lost in cyclostomes, and the other gave rise to Mb and Cygb gene lineages (*right branch*). However, in cyclostomes Mb was secondarily lost. During their evolution both lineages independently acquired similar physiological function of oxygen transportation (denoted by *orange* line, one in the Hb lineages of gnathstomes at about 500 MYA and the other in cytoglobin lineage (referred as Hb) of cyclostomes at around similar time. According to this model, cyclostome Hbs are not orthologous to gnathostome Hbs but are orthologous to gnathostome Cygbs. Adapted from Hoffmann et al. (2010a).

The phylogenetic relationship and evolution of vertebrate globins still remains controversial but the recent study performed by Hoffmann et al. (2010a) grouped them into four main clades: (i) cyclostome Hb + Cygb, (ii) Mb + GbE, (iii) GbY, and (iv) gnathostome Hb (Figure 1A). These different globin genes evolved and diversified further to adapt to the atmospheric oxygen levels, and requirements for oxygen in different species and at different developmental stages. Each of these then acquired specialised functions such as oxygen transportation or storage, nitric oxide metabolism or detoxification, other enzymatic function, collagen synthesis or electron transfer (Burmester et al., 2004, Fago et al., 2004, Burmester et al., 2007, Trandafir et al., 2007). According to Hoffmann et al. (2010a) Hb in gnathostomes and cyclostomes evolved independently and acquired similar function (Figure 1B).

## Haemoglobin

Among globin families, the oxygen-carrying protein Hb has been of extreme interest for many decades. It was first discovered by Hünefeld in 1840 and since then it has been among the first few proteins to be studied extensively to understand the crystallographic structure, structure-function relationship of proteins, structural transitions between conformers, ligand binding, allosteric interactions, relation between mutation and protein change, regulation and globin gene evolution (Perutz, 1984, Berenbrink, 2006, de Souza and Bonilla-Rodriguez, 2007).

Haemoglobin originated as an iron-complexed protein at about the time when life originated on earth. Since it has been evolving for approximately 4 billion years, selective pressures and changes in atmospheric oxygen levels have led to diversity of specialised forms and functions of Hb across all kingdoms. They are found in all groups of organisms, including prokaryotes, fungi, plants and animals and their functions range from catabolic metabolism in bacteria, yeast and worms to oxygen transportation in vertebrates (reviewed in Hardison, 1998).

### Haemoglobin in vertebrates

Haemoglobin is found in the blood of almost all vertebrates carrying oxygen from lungs/gills to respiring tissues in need and then carrying carbon dioxide from these tissues back to lungs/gills. The name *haemoglobin* is derived from the words *haem* and

*globin*, because Hb is made up of globular protein subunits, each of which embeds a haem pocket that contains one iron atom. This iron atom can reversibly and cooperatively bind an oxygen molecule through ion-induced dipole forces. In gnathostomes, a Hb molecule is made up of four such subunits, two alpha (α)- and two beta (β)- globins. However, in cyclostomes (represented by lampreys and hagfish) Hbs appear as monomers in the oxygenated state but self-associate into dimers or tetramers upon deoxygenation (Briehl, 1963, Rumen and Love, 1963, Li and Riggs, 1970).

As well as Hb structures, the cooperative binding properties of oxygen are also different in gnathostomes and cyclostomes. In the $\alpha_2\beta_2$ Hb tetramers of most extant gnathostomes, the cooperative binding of oxygen is a result of oxygenation-linked transition in quaternary structure (Coates, 1975, Goodman et al., 1975, Goodman et al., 1988), whereas, in the Hbs of extant cyclostomes, it is a result of oxygenation-linked dissociation of multimers into ligated monomers (Briehl, 1963, Rumen and Love, 1963, Li and Riggs, 1970).

The evolution of erythroid Hbs of cyclostomes and gnathostomes has not been clear. It was long thought that cyclostome Hbs were orthologous to gnathostome Hbs and shared a common ancestor (Goodman et al., 1988). Only recently has molecular phylogeny of all members of the globin superfamily clarified that they are paralogous and evolved independently with similar biochemical properties (Figure 1B) (Katoh and Miyata, 2002, Hoffmann et al., 2010a).

In the gnathostome ancestor the α- and β-globin genes evolved by a tandem duplication of a single ancestral primordial globin gene approximately 450-500 MYA (Czelusniak et al., 1982, Goodman et al., 1987, Lanfranchi et al., 1994). These genes underwent repeated round of duplications and divergences (including lineage-specific gene loss and gain) to form a repertoire of α- and β-globin genes located together in teleost fish and frogs but separated into distinct clusters in amniotes. The human Hb genes in particular, have attracted more interest in this area of research than other species, with the aim of gaining a better understanding of Hb synthesis, functions and associated diseases. Since many Hb related hereditary diseases, termed haemoglobinopathies in humans, are caused by mutations in the genes encoding for Hb, most studies have focused on structures and regulatory pathways of the α- and β-globin genes and their evolution in humans and model mammalian species.

**Human Hb: Synthesis, Regulation and Diseases**

Throughout human life, different isoforms of Hb are produced to meet specific requirement for oxygen (Weatherall, 1991). For example, in the human embryo the Hb molecule is made up of two embryonic α- and β-like globin chains ($\zeta_2\varepsilon_2$ called Hb Gower 1) while in the foetus it is made up of two foetal/adult α-like and two foetal β-like globin chains ($\alpha_2\gamma_2$ called HbF), and in an adult it is made up of two adult α- and β-like globin chains ($\alpha_2\beta_2$ called HbA). This results in expression from two distinct clusters of α- and β-globin genes that are co-ordinately regulated to synthesize different genes from these clusters at different developmental stages and tissues.

The human α-globin cluster (5'- *HBZ-T1, HBZ-T2ps, HBK, HBA-T1ps, HBA-T2, HBA-T3, HBQ* -3') is located on chromosome 16 in a constitutively active chromatin and CpG rich domain, which is a dominant determination of its regulation (Orkin, 1978, Lauer et al., 1980, Proudfoot and Maniatis, 1980, Hardison et al., 1986). The β-globin cluster (5'- *HBE, HBG-T1, HBG-T2, HBHps, HBD, HBB* -3'), however, is located in an A+T rich genomic region on chromosome 11 (Flavell et al., 1978, Bernard et al., 1979, Fritsch et al., 1980). The loci designated 'ps' at the end of the gene denote pseudogenes that are no longer transcriptionally active and functional due to mutations, and may be evolutionary remnants of once active globin genes (Weatherall, 1991). The genes in these clusters are arranged in their order of expression: *HBZ* and *HBE* are expressed in an embryo followed by *HBG* in the foetus and then *HBA*, *HBD* and *HBB* in late embryogenesis and postnatal life (Figure 2). Changes in their expression coincide with changes in the anatomic site of haematopoiesis. For example, embryonic globin genes are expressed during primitive erythropoiesis in the egg yolk, foetal globin genes are expressed in the liver of the foetus, followed by adult globin genes in the bone marrow during definitive erythropoiesis. Similar expression patterns are also seen in other amniotes.

The specialised forms of Hb at each developmental stage have different affinities for oxygen that are advantageous to that stage. Early expressed Hbs, embryonic and foetal for instance, lack cooperative oxygen binding and have higher oxygen affinity than the adult Hb. This enables the embryo/foetus living in a hypoxic environment to extract more oxygen with greater affinity from the maternal blood in the placenta (Merlet-Benichou, 1975). As the gas exchange structures develop and placental barriers are

reduced, these high oxygen affinity Hbs are then slowly replaced by lower affinity Hbs with greater cooperatively (Brittain et al., 1997).

Although a Hb molecule may look simple, its whole coordination and regulatory pathway is so complex that researchers have not yet been able to fully understand it, despite decades of research. 'Globin switching' in different ontogeny and tissues are regulated mainly by many transcription binding factors that bind to distal *cis*-regulatory regions, enhancers and to local promoters associated with each gene.

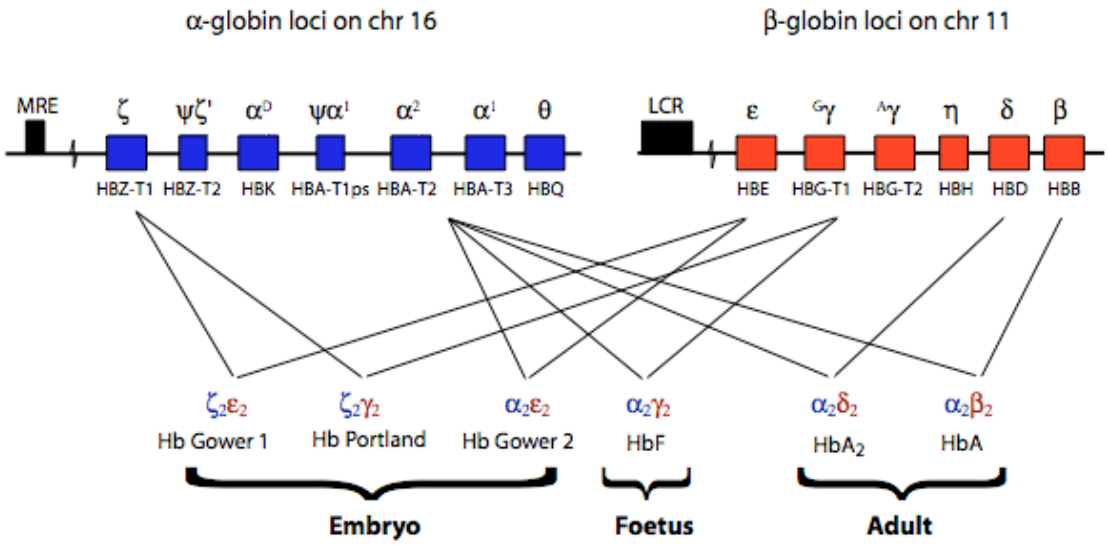



**Figure 2: Globin switching of Hb synthesis in humans**

The genes in the α- and β-globin clusters (located on different regions) are developmentally regulated by MRE (also known as HS-40) and LCR respectively, to form different Hb isoforms in different stages and tissues. The α-globin genes are in *blue*, the β-globin genes are in *red* and the regulatory regions are in *black*. Their Greek names are on top, whilst their standard nomenclature names are on bottom of genes.

The major regulatory element (MRE) and locus control region (LCR) are the major *cis*-regulatory elements thought to be involved in the timing and erythroid cell lineage- and developmental stage-specific expression of α- and β-globin genes, respectively. These regions possess many DNase hypersensitive sites (HSs) that contain binding sites for many transcription factors such as GATA-1, AP1/NF-E2, EKLF, SP-1 and enhancers that control the high level, copy number-dependent, position-independent expression of associated globin gene in transgenic mice (Grosveld et al., 1987b, Higgs et al., 1990,

Hardison et al., 1997, Hardison, 1998). The MRE (also known as HS-40) is located approximately 40 kb upstream of the α-globin locus, in an intron of the adjacent gene (*C16orf35*) and contains a single erythroid-specific HS spanning 350 bp (Higgs et al., 1990, Vyas et al., 1995). On the other hand, LCR is located about 10 kb upstream of β-globin locus and is about 20-25 kb long, comprising five HSs (HS 1-5) (Forrester et al., 1986, Grosveld et al., 1987b, Stamatoyannopoulos, 2005). Thus, Hb synthesis uses a highly coordinated mechanism to control developmental stage-specific and tissue-specific expression of globin genes.

Defects or mutations anywhere in the α- or β- globin genes, regulatory regions, promoters or transcription factors can affect the structure, production and function of Hb, thereby resulting in haemoglobinopathies. The most common ones include sickle-cell disease and thalassaemia that involve imbalanced production of either α- or β-globin proteins resulting in the underproduction of normal and sometimes abnormal Hbs. The effects of haemoglobinopathies are decreased oxygen binding (causing shortness of breath), anaemia, fatigue, altered craniofacial complexes, bone pains, growth failure, organ damage and at times early death (Weatherall, 1991, Bassimitci et al., 1996, Garner et al., 2003). Although human patients suffer from anaemia and thalassaemia due to mutations in α- or β-globin genes, the mutations are maintained in the population through natural selection by indirectly protecting them against malaria (Allen et al., 1997, Fucharoen and Winichagoon, 1997, Mockenhaupt et al., 2004).

Since there are about 1000 haemoglobinopathies (some major while others minor) identified in humans (Wells and Brennan, 1992, Labie and Elion, 1996, Wells, 1999), there is a huge medical, clinical and commercial interest in diagnosing and treating these inherited diseases to increase the lifespan and the quality of life in patients. Comparative genomic research into Hb gene structures and regulatory pathways, and the effects of mutations in different genomic regions in humans and other species is critical for understanding and testing for these genetic diseases, as it can pinpoint the molecular basis of Hb disorders and functions in humans.

**Evolution of Hb repertoire in gnathostomes**

Many studies have developed models to explain the evolution of the α- and β-globin genes in vertebrates, that is, how a single primordial globin gene duplicated 450-500

MYA into paralogous α- and β-globin genes located close together, became separated into two distinct clusters located on different chromosomes as seen in humans. These models have either been developed further or refuted as information from additional species has become available. In this section, I will review the organisation of α- and β-globin in vertebrates and the theories put forward to explain their evolution in amniotes.

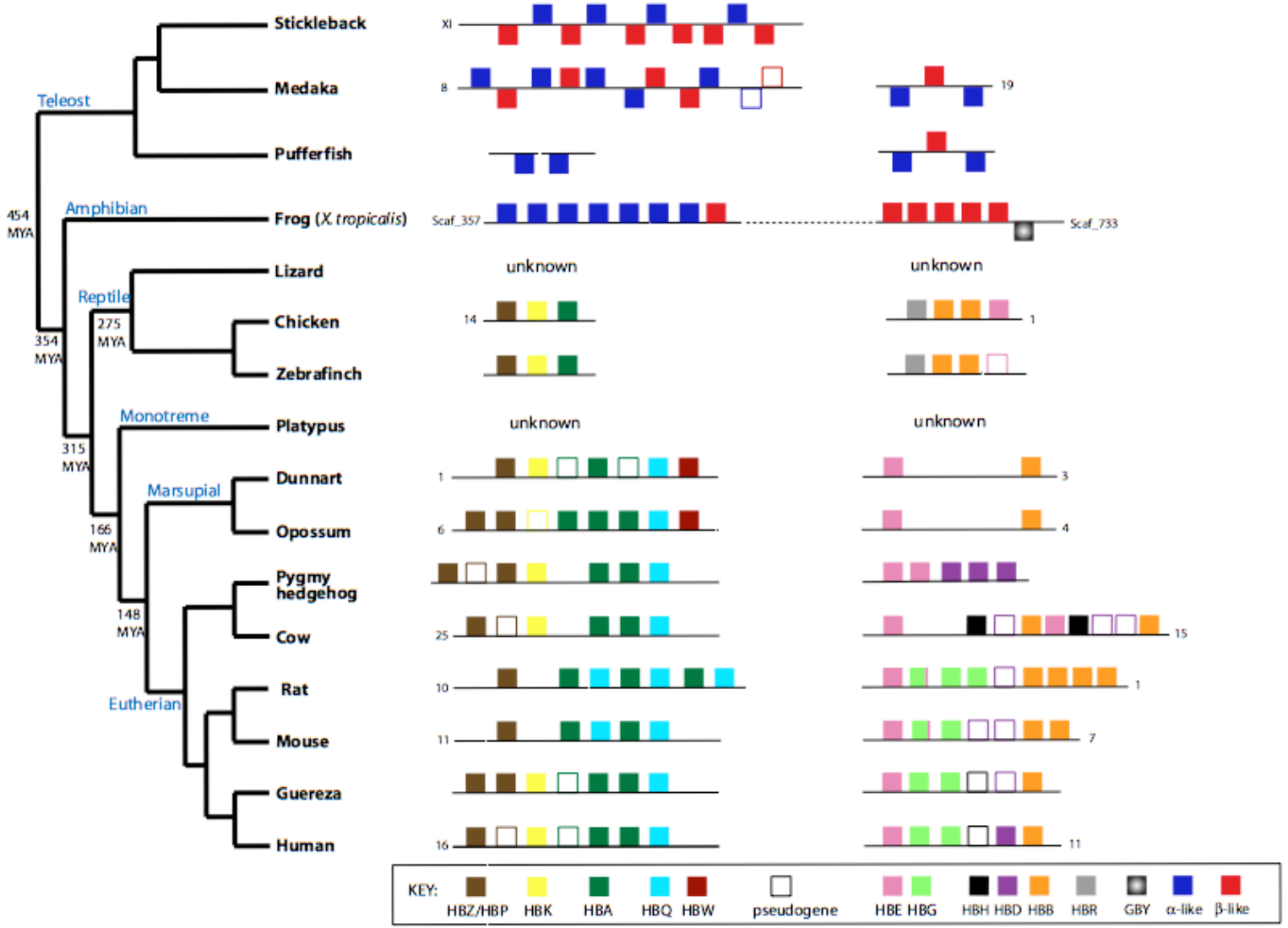
In the teleost lineage, both genes duplicated *in blocs* and specialised into 'early-expressed' in larva ($\alpha_L$ and $\beta_L$) and 'late-expressed' in adult ($\alpha_A$ and $\beta_A$) globin genes, which further duplicated and diverged in each of teleost lineages. These genes are located in a head-to-head orientation in respect of transcriptional polarity, meaning they are transcribed in opposite directions in the general order $^{3'}\alpha_L{}^{5'}$-$^{5'}\beta_L{}^{3'}$-$^{3'}\alpha_A{}^{5'}$-$^{5'}\beta_A$-$^{3'}$ but with different copy numbers of each gene in different teleost species (Figure 3) (Brownlie et al., 2003, Gillemans et al., 2003, Maruyama et al., 2004, Quinn et al., 2010). Some teleosts such as zebrafinch (*Danio rerio*), pufferfish (*Fugu rubripes*), medaka (*Oryzias latipes*) and Atlantic salmon (*Salmo salar*) exhibit two clusters of intermixed α- and β-globin genes located on different chromosomes. However, in the stickleback (*Gasterosteus aculeatus*) there is only one α-β globin cluster (Quinn et al., 2010). This suggests that the presence of two clusters with different flanking genes in other teleosts is a result of whole genome duplication (WGD) at the base of teleost radiation (Kassahn et al., 2009), but the second cluster was lost in stickleback (Quinn et al., 2010). In Atlantic salmon, the second cluster was also lost after the teleost-specific WGD but before the salmonid-specific WGD, which produced the duplicated copies of the first cluster (with same flanking genes) seen today (Figure 3) (Quinn et al., 2010).

However, in the frog lineage there was another set of independent duplications of the ancestral α- and β-globin genes to also result in 'early-expressed' in embryo ($\alpha^E$ and $\beta^E$) and 'late-expressed' in adult ($\alpha^A$ and $\beta^A$) globin genes followed by further tandem duplications and divergences. Unlike teleosts, the α- and β-globin genes of the African clawed frog *Xenopus laevis* are not intermixed but located on either sides of the cluster arranged in the general order 5'-$\alpha^E$-$\alpha^A$-$\beta^A$-$\beta^E$-3' and with the same transcriptional polarity, that is, all genes are located in a head-to-tail orientation that permits transcription in same direction (Patient et al., 1980, Hosbach et al., 1983). This is also seen in the Western clawed frog *Xenopus tropicalis*, but it is uncertain whether they contain one or two clusters of globin genes, as its current genome assembly shown on

the Ensembl genome browser (v 4.1, August 2005 on Ensembl http://www.ensembl.org/index.html) shows two different scaffolds containing globin genes. This was also documented by Fuchs et al. (2006) but they speculated that these two scaffolds are actually linked, thus showing the similar arrangement as *X. laevis* (refer to Figure 3 for details into their cluster). However, a more complete assembly and additional analyses, such as mapping of these scaffolds, are needed to resolve this issue. Previous studies in *X. laevis* clearly showed the presence of two distinct α-β clusters on different chromosomes but was determined to be a result of tetraploidisation of an *X. tropicalis*-like ancestor, perhaps 10-30 MYA (Bisbee et al., 1977, Jeffreys et al., 1980, Hosbach et al., 1983). In *X. laevis*, globin genes are differentially expressed at stages and in tissues similar to other gnathostomes including human (Banville et al., 1983, Banville and Williams, 1985b).

In addition to the α-β globin cluster in *X. tropicalis*, another globin-related gene *GBY* is located 3' to this cluster (Fuchs et al., 2006). The discovery of *GBY* in frogs provided the first case of two different globin family members located together (in the arrangement α-β-*GBY*) that do not functionally interact. In fact, the function of *GBY* is unknown at this stage although it is transcribed into a functional polypeptide chain that has the hallmarks of a respiratory protein (Fuchs et al., 2006).

**Figure 3 (following page): Genomic structures of α- and β-globin gene clusters in teleosts and tetrapods.** The orientation of the clusters are drawn from 5' (on the left) to 3' (on the right) but not drawn to scale. The amniote globin genes are referred according to their standard nomenclatures, however, in teleosts and frogs, their globin genes are referred by generic 'α-globin' and 'β-globin' as they are less well characterised. Information on these clusters were taken from several sources: mammalian α (Hoffmann et al., 2008), mammalian β (Opazo et al., 2008a), marsupial α (Cooper et al., 2006), cow β (Schimenti and Duncan, 1985), bird α and β (Alev et al., 2009), frog α-β (Fuchs et al., 2006, Hardison, 2008), pufferfish (Gillemans et al., 2003), medaka (Maruyama et al., 2002), salmon (Quinn et al., 2010) and stickleback (Broad/gasAcul Assembly Feb 2006 from UCSC browser http://genome.ucsc.edu/). Numbers inside the phylogeny show dates of divergence in million years ago (MYA) (Bininda-Emonds et al., 2007, Hedges and Vidal, 2009, Shedlock and Edwards, 2009) and numbers beside gene clusters show chromosomal locations.


Stickleback
Medaka
Pufferfish
Frog (X. tropicalis)
Lizard
Chicken
Zebrafinch
Platypus
Dunnart
Opossum
Pygmy hedgehog
Cow
Rat
Mouse
Guereza
Human
Teleost
Amphibian
Reptile
Monotreme
Marsupial
Eutherian
454 MYA
354 MYA
315 MYA
275 MYA
166 MYA
148 MYA
XI
8
19
Scaf_357
Scaf_733
unknown
14
1
6
25
10
11
16
3
4
15
7
KEY:
HBZ/HBP
HBK
HBA
HBQ
HBW
pseudogene
HBE
HBG
HBH
HBD
HBB
HBR
GBY
α-like
β-like

Before the amniote evolution (>315 MYA), the α- and β-clusters separated onto different chromosomes, thereby, becoming unlinked in amniotes. Jeffreys et al. (1980) originally proposed two models explaining the separation of α- and β-globin clusters in amniotes. One model hypothesized that both clusters became unlinked by translocation between the α- and β-globin genes, possibly by chromosomal rearrangements after the amphibian-amniote divergence (Pisano et al., 2003). The alternative model hypothesized that *in trans* duplication (chromosomal or genome) of the entire α-β globin cluster resulted in two clusters (proto α1-β1 and proto α2-β2), which then underwent differential silencing in amniotes (Jeffreys et al., 1980). For example, the proto-β1 was silenced and left proto-α1 as the α-globin cluster and similarly, the proto-α2 was silenced and left proto-β2 as the β-globin cluster, therefore unlinking both clusters in amniotes. However, this model leaves the possibility that some evolutionary remnants of inactivated β1 or α2 would be present near contemporary amniote α- or β-globin clusters, respectively.

*Translocation / chromosomal rearrangement theory*

For many years, the translocation / chromosomal rearrangement theory of Jeffreys et al. (1980) had been supported by the presence of two distinct clusters each containing a set of differentially expressed α- or β-globin genes located on separate chromosomes in amniotes. After the separation of the clusters, the genes contained in them duplicated and diverged further in different amniote lineages.

The amniote ancestor contained three α-like globin genes in their α-globin cluster in the order 5'-*HBP-HBK-HBA*-3'. The *HBP* and *HBA* are orthologues of the frog 'early-expressed' (embryonic $\alpha^{E}$) and 'late-expressed' (adult $\alpha^{A}$) α-globin genes, respectively. The *HBK* gene evolved by duplication of either *HBP* (Hoffmann and Storz, 2007, Hoffmann et al., 2010b) or *HBA* (Cooper et al., 2006), but the order and timing of this duplication is still not clear. In birds (represented by chicken *Gallus gallus* and zebrafinch *Taeniopygia guttata*) all three genes are present in the same order and encode functional proteins (Engel and Dodgson, 1980, Dodgson et al., 1981, Alev et al., 2009). However, in the common ancestor of marsupials and eutherians the *HBP* duplicated once and *HBA* duplicated three times to result in a seven-gene α-globin cluster 5'- *HBZ-T1, HBZ-T2, HBK, HBA-T1, HBA-T2, HBA-T3, HBQ*-3' in marsupials, similar to those of humans (Cooper et al., 2006). However, the sequences of two

paralogous copies of embryonic *HBZ* and three paralogous copies of adult *HBA* globin genes have been homogenised by concerted evolution such as inter-paralogue gene conversion or by unequal crossing-over (Cooper et al., 2006, Hoffmann et al., 2008). Furthermore, subsequent gene duplications, losses and/or pseudogenisations in some eutherian species resulted in different copy numbers of each gene (Figure 3).

In the therian α-globin clusters, there are two genes that are expressed in some marsupials and eutherians, including humans, but their functions are still unknown. The *HBK* is transcribed in cats, dogs, pigs and cows and even in humans, but no *HBK* containing Hb has been detected (Goh et al., 2005, Cooper et al., 2006). Computational analysis suggests that *HBK* is evolving under purifying selection and is therefore, likely to be functional (Cooper et al., 2006). In humans, *HBK* has an expression pattern similar to that of its reptile orthologue, that is, it is expressed in embryonic, foetal and adult stages of development. The only difference is that in reptiles it forms a functional Hb molecule (Hiebl et al., 1986, Abbasi et al., 1988, Rucknagel and Braunitzer, 1988, Fushitani et al., 1996) whereas in mammals it does not form a functional Hb molecule, and its function remains unknown. Similarly, *HBQ* is expressed in some eutherian mammals, including humans, in both embryonic and adult stages of development (Leung et al., 1987, Hsu et al., 1988, Albitar et al., 1989, Albitar et al., 1992). It is expressed also in marsupial embryos (Cooper et al., 2005), but no *HBQ*-containing Hb has been detected. More research is needed into both *HBK* and *HBQ* globin genes to gain insight into their possible roles and importance to humans.

Unlike the amniote α-globin genes, the evolution of the β-globin genes in amniotes and the arrangement of genes in the ancestral β-globin cluster are not clear, partially because the bird and mammalian β-like globin genes differ in sequences, numbers, expression patterns and regulatory elements. The bird β-globin clusters contain four genes 5'- *HBR, HBB-T1, HBB-T2, HBE* -3' (*HBE* is a pseudogene in zebra finch) that are not arranged in their order of expression; the other two genes (*HBR* and *HBE*) are expressed in the early stages of development while the genes sandwiched in between are expressed in later stages (Bruns and Ingram, 1973, Reitman et al., 1993a, Mason et al., 1995, Alev et al., 2009).

The β-globin clusters of therian mammals, on the other hand, are arranged in their order of expression; the early-expressed genes on 5′ end and late-expressed genes on 3′ end,

similar to the amniote α-like globin genes. Since marsupials have two β-like globin genes, *HBE* (embryonic) at 5' end and *HBB* (adult) at 3' end (Figure 3) in their cluster, which are also present in eutherians, the marsupial β-globin cluster represents the ancestral therian state (Koop and Goodman, 1988, Cooper and Hope, 1993, Cooper et al., 1996, De Leo et al., 2005). After the divergence of marsupials from eutherians but prior to the eutherian radiation, the early-expressed *HBE* duplicated twice and the adult-expressed *HBB* duplicated once to form the ancestral eutherian β-globin cluster (5'-*HBE, HBG-T1, HBG-T2, HBH, HBD, HBB* -3') (Goodman et al., 1974, Goodman et al., 1984, Koop and Goodman, 1988, Weatherall, 1991, Hardison and Miller, 1993, Opazo et al., 2008a). These then evolved in different eutherian lineages into the complex β-globin cluster (Figure 3) through different duplication events, gene gain and loss, gene conversion and/or unequal crossing-over. For example, in the goat *Capra hirus,* the β-globin cluster contains a triplication of a set of four genes, 5'- *HBE, HBH, HBD, HBB* -3' (Townes et al., 1984) and similar *en bloc* duplication occurred in the lineage that includes the cow *Bos taurus* (Schimenti and Duncan, 1985) and sheep *Ovis aries* (Garner and Lingrel, 1988, Garner and Lingrel, 1989).

The bird β-like globin genes show fewer similarities than expected for orthologous genes to the mammalian β-like globin genes, and their regulatory regions show no similarity at all to those of other mammalian β-globins (Reitman et al., 1993a). Each gene in the bird β-globin cluster is equally similar to all mammalian β-like globin genes. Unlike mammals, the chicken cluster lacks a strong regulatory region (LCR) upstream. Instead, there is a strong tissue-specific enhancer located between the *HBB-T2* and *HBE* to control tissue-specific expression of β-globin genes (Choi and Engel, 1986, Hesse et al., 1986, Reitman et al., 1993a). These differences including copy numbers and expression patterns suggest that the bird and mammal β-globin genes are not orthologous, but resulted from independent duplications of an ancestral β in the bird and mammalian lineages (Reitman et al., 1993a).

*In trans duplication of α-β loci*

About a decade ago, the unexpected discovery of the third marsupial β-globin gene called *HBW* (ω-globin) adjacent (3') to the α-globin cluster revisited the evolution model. This "orphan" *HBW* is not a part of the main β-globin cluster (Wheeler et al.,

2001, Wheeler et al., 2004) and is expressed in marsupial neonates just before birth and in early pouch young development, and forms a function haemoglobin molecule with α-globin chains (Holland and Gooley, 1997, Holland et al., 1998). Sequence comparisons showed that the marsupial *HBW* was more similar to chicken β-like globin genes than to the major cluster of β-like globin genes in marsupials, or to those of other mammals (Wheeler et al., 2001, Wheeler et al., 2004).

The revolutionary discovery of *HBW* was important because it refuted the translocation/chromosomal rearrangement theory and instead provided support for the alternate hypothesis postulated by Jeffreys et al. (1980) suggesting that β-globin evolution in birds and mammals represents more ancient genome or chromosomal (*in trans*) duplications within their common ancestor. The modified hypothesis proposed that an ancestral α-β globin duplication followed by differential loss resulted in two independent unlinked gene clusters prior to bird-mammal divergence. One cluster diverged to form the β cluster in birds, but became redundant and inactive, and was finally deleted in mammals, except for the marsupial *HBW*, which is seen as the last mammalian relic of this cluster. The other cluster diverged to form the β cluster in mammals, including marsupials (Wheeler et al., 2001, Wheeler et al., 2004). Thus, according to this model, the bird β-globin cluster is paralogous to the mammalian β-globin cluster, contracting the previous claims that both clusters were orthologous.

This model explained why chicken β-like globin genes did not have high sequence similarities to mammalian β-like globin genes (Reitman et al., 1993a, Hardison, 1998) but failed to explain why the regions spanning the duplicated globin clusters are so different, i.e the bird and therian α-globin clusters reside in a G+C rich domain that contains a CpG island, but their β-globin clusters reside in a A+T rich domain (reviewed in Hardison, 1998).

## Importance of globin studies in non-avian reptiles and monotremes

The *in trans* duplication model proposed by Wheeler et al. (2001, 2004) is not well supported at two places in amniote phylogeny, one at the stem of mammalian evolution where monotremes diverged from therians about 166 MYA and the other at the stem of amniote evolution where reptiles diverged from mammals about 315 MYA (Figure 3). A great deal of information is lacking from non-avian reptiles and monotremes that

would provide valuable comparisons with bird and therian globin genes and would validate claims made for the origin of β-globin genes in birds. It would also assist in constructing the ancestral globin clusters of amniotes and mammals.

To date, there have been no molecular studies of entire globin clusters in non-avian reptiles and monotremes. Previous studies in non-avian reptiles have isolated and characterized some haemoglobin proteins (both embryonic and adult) but they contained amino acid sequences of adult α- and β-globin proteins only and not embryonic globins (Gorr et al., 1998 and references therein). Thus, we do not know if non-avian reptiles possess any embryonic globin genes and if any, how are they organised. Research is therefore needed into globin genes in non-avian reptiles, to gain insight into the gene structure of their clusters, construct ancestral globin clusters at the stem of reptile radiation and to chart the course of globin gene evolution in tetrapods.

Similarly, early work on monotreme globins consisted of the isolation of adult α- and β-globin proteins in the Australian echidna (*Tachyglossus aculeatus*) and platypus (*Ornithorhynchus anatinus*) (Whittaker et al., 1972, Cooper et al., 1973, Thompson et al., 1973, Whittaker and Thompson, 1974, Whittaker and Thompson, 1975, Lee et al., 1999). However, there have been no reports of the existence of any embryonic globin genes in either species. This leaves many questions unanswered about the gene structure and organization of α- and β-globin clusters in monotremes that could help construct ancestral mammalian globin clusters and discover the origin of specialised mammalian globins. It is critical to know whether monotreme α- and β-globin clusters are similar to those in therian mammals or birds, or whether there are globins specific to monotremes that reflect their unique lineage and lifestyle.

The importance of non-avian reptiles and monotremes to our understanding of the eutherian (indeed the human) genome has been recently appreciated sufficiently to sequence the genomes of first squamate reptilian species, *Anolis carolinensis* (green anole lizard) at the depth of 6.3 times and a monotreme species, platypus at the depth of 6 times. The bacterial artificial chromosome (BAC) library, tissues, metaphase preparations and other materials are readily available for the platypus, all of which are useful resources for globin studies. However, these materials are not readily available in Australia for the green anole, but another model lizard species, the Australian central bearded dragon *Pogona vitticeps* has no such restrictions. Combining research on globin

genes in both lizard species would provide constructive comparisons between reptiles and with other amniotes.

## The present study; aims and chapter précis

Few data are available concerning α- and β-globin genes in monotreme mammals and non-avian reptiles. Filling in these phylogenetic gaps will provide a much clearer picture of globin gene evolution in amniotes, including their functions and regulatory elements. The draft genome assemblies of platypus, an Australian iconic mammal, and the American green anole lizard, and the availability of the BAC library of the Australian bearded dragon lizard, provide opportunities to analyse the entire α- and β-globin regions in these species. What genes are present in their clusters and are they conserved with other amniotes? Is there any *HBW* present beside the α-globin cluster? Do they contain any embryonic globin genes? Will they reveal new genes or regulatory regions? Will a more detailed analysis of globin genes and flanking sequences in monotremes and lizards help construct ancestral clusters at the stem of amniote evolution? I seek to answer these questions in this thesis.

In the present study, α- and β-globin genes and their flanking genes will be investigated at the molecular level in the platypus, and in two squamate lizards; the bearded dragon and green anole, which represent different lizard families in two major cohorts so they will allow valid comparisons between them and to other reptiles.

Using the above study species, the main aim of this study was to understand how the complex α- and β-globin gene clusters and regulatory regions evolved in amniotes. To address these questions, I present here the results of my research in the format of two first-author research publications (Chapter 2 and 4B), one first-author review article (Chapter 3) and two draft manuscripts (Chapter 4A and 5).

**Aim 1:** To characterise α- and β-globin genes in the platypus, deduce the content and arrangement of the ancestral globin clusters at the stem of mammalian radiation, and study globin gene evolution in amniotes.

Chapter 2 provides a comprehensive research study and methodology into isolating and characterising the platypus globin genes and constructing the

ancestral mammalian globin cluster. Globin gene evolution was revisited as my results contradicted the previous model (Wheeler et al., 2001, Wheeler et al., 2004). I proposed a new transpositional model for globin gene evolution in amniotes, which has won wide acceptance.

**Aim 2:** To understand the mechanisms behind the transposition of an original β-globin gene into a new site, and the effect of such transposition.

Chapter 3 reviews the literature on globin genes in vertebrates and discusses how and why a transposition event could have occurred, and what effect it might have had on globin regulation in amniotes.

**Aim 3:** To characterise α- and β-globin gene clusters in lizards to further test the transposition hypothesis.

Chapter 4 is divided into two parts. The first part (4A) contains a bioinformatic analysis of the green anole α- and β-globin sequences, and discussion of their evolution in reptiles, which led me to research into the Australian dragon lizard. This section was not included with the below manuscript because another independent research group published a similar analysis of the same data before my paper was accepted. I present this drafted manuscript to demonstrate my approach and mastery of bioinformatic analysis. The second part (4B) is a research publication that, in the absence nucleotide sequence data, describes the use of molecular biology approaches to isolate and map α- and β-globin clusters in the dragon lizard. My results provide further support for the transpositional model.

**Aim 4:** To discover major regulatory regions of the platypus α- and β-globin clusters and to study their evolution in amniotes.

Chapter 5 (paper in preparation) describes a number of techniques I used to predict the major regulatory region (MRE) of the platypus α-globin cluster and the locus control region (LCR) of the platypus β-globin cluster. My results showed that the platypus MRE is conserved to some extent in sequence, transcription factor binding motifs and its ability to enhance gene expression amongst jawed vertebrates, reflecting the α-globin loci's ancient origin. However, my failure to uncover a platypus regulatory element

conserved with either therian mammals or chickens using many exhaustive searches, suggests that the regulatory regions of the monotreme β-globin cluster have either diverged significantly or evolved independently along with their genes. In this chapter I discuss possible events that gave rise to modern regulatory regions of amniote β-globin cluster.

In the final chapter (Chapter 6) I integrate my results from above chapters and from a publication to which I contributed (shown in Appendix 1) to discuss the significance of my findings with respect to other published studies, discuss their implications and suggest future directions. Appendix 2 lists oral presentations and posters presented at local and international conferences.

# CHAPTER 2: PLATYPUS GLOBIN GENES AND FLANKING LOCI SUGGEST A NEW INSERTIONAL MODEL FOR β-GLOBIN EVOLUTION IN BIRDS AND MAMMALS

***Authors:***

**Patel, V.S.,** Cooper, S.J.B., Deakin, J.E., Fulton, B., Graves, T., Warren, W.C., Wilson, R.K., and Graves, J.A.M.

***Current status of paper (circle as appropriate):***

Planned/ In Preparation/ Submitted/ Under Revision/ Accepted/ Published

***Date paper accepted for publication or anticipated date of publication:***

25 July 2008

***Name of journal:***

BMC Biology **6**: 34

***Extent to which research is your own:***

I designed and performed all experiments and analysed data (except phylogenetic analyses which was performed by authors SJBC, and BAC sequencing which was done by BF, TG, WCW and RKW).

***Your contribution to writing the paper:***

I wrote the draft of the manuscript (except the section on phylogenetic analyses, which was drafted by SJBC), prepared figures and incorporated suggestions made by co-authors. I corresponded with the journal editor and responded to the referees' comments

***Comments:***

In this manuscript, I have provided comprehensive research and analyses of platypus globin genes and their flanking regions. This was my major experimental work; it took two years and much effort to perform and trouble-shoot various experiments, and to analyse and present the data in this manuscript.

BMC Biology

BioMed Central

Research article

Open Access

# Platypus globin genes and flanking loci suggest a new insertional model for beta-globin evolution in birds and mammals

Vidushi S Patel*[1], Steven JB Cooper[2,3], Janine E Deakin[1], Bob Fulton[4], Tina Graves[4], Wesley C Warren[4], Richard K Wilson[4] and Jennifer AM Graves[1]

Address: [1]The ARC Centre for Kangaroo Genomics, Research School of Biological Sciences, The Australian National University, Canberra, ACT 2601, Australia, [2]Australian Centre for Evolutionary Biology and Biodiversity, The University of Adelaide, Adelaide, SA 5005, Australia, [3]Evolutionary Biology Unit, South Australian Museum, Adelaide, SA 5000, Australia and [4]Genome Sequencing Center, Washington University School of Medicine, St Louis, Missouri 63108, USA

Email: Vidushi S Patel* - vidushi.patel@anu.edu.au; Steven JB Cooper - Cooper.Steve@saugov.sa.gov.au; Janine E Deakin - janine.deakin@anu.edu.au; Bob Fulton - bfulton@watson.wustl.edu; Tina Graves - tgraves@watson.wustl.edu; Wesley C Warren - wwarren@watson.wustl.edu; Richard K Wilson - rwilson@watson.wustl.edu; Jennifer AM Graves - jenny.graves@anu.edu.au

* Corresponding author

Published: 25 July 2008

*BMC Biology* 2008, **6**:34 doi:10.1186/1741-7007-6-34

Received: 28 May 2008
Accepted: 25 July 2008

This article is available from: http://www.biomedcentral.com/1741-7007/6/34

© 2008 Patel et al; licensee BioMed Central Ltd.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

**Background:** Vertebrate alpha (α)- and beta (β)-globin gene families exemplify the way in which genomes evolve to produce functional complexity. From tandem duplication of a single globin locus, the α- and β-globin clusters expanded, and then were separated onto different chromosomes. The previous finding of a fossil β-globin gene (ω) in the marsupial α-cluster, however, suggested that duplication of the α-β cluster onto two chromosomes, followed by lineage-specific gene loss and duplication, produced paralogous α- and β-globin clusters in birds and mammals. Here we analyse genomic data from an egg-laying monotreme mammal, the platypus (*Ornithorhynchus anatinus*), to explore haemoglobin evolution at the stem of the mammalian radiation.

**Results:** The platypus α-globin cluster (chromosome 21) contains embryonic and adult α- globin genes, a β-like ω-globin gene, and the *GBY* globin gene with homology to cytoglobin, arranged as 5'-ζ-ζ'-$\alpha^D$-$\alpha^3$-$\alpha^2$-$\alpha^1$-ω-*GBY*-3'. The platypus β-globin cluster (chromosome 2) contains single embryonic and adult globin genes arranged as 5'-ε-β-3'. Surprisingly, all of these globin genes were expressed in some adult tissues. Comparison of flanking sequences revealed that all jawed vertebrate α-globin clusters are flanked by *MPG-C16orf35* and *LUC7L*, whereas all bird and mammal β-globin clusters are embedded in olfactory genes. Thus, the mammalian α- and β-globin clusters are orthologous to the bird α- and β-globin clusters respectively.

**Conclusion:** We propose that α- and β-globin clusters evolved from an ancient *MPG-C16orf35*-α-β-*GBY-LUC7L* arrangement 410 million years ago. A copy of the original β (represented by ω in marsupials and monotremes) was inserted into an array of olfactory genes before the amniote radiation (>315 million years ago), then duplicated and diverged to form orthologous clusters of β-globin genes with different expression profiles in different lineages.



**Publication (Patel et al. 2008)**

## Background

The evolution of the vertebrate globin superfamily has been extensively studied for many decades by comparing the structure and function of members of the gene families. These are principally haemoglobin, myoglobin, cytoglobin and neuroglobin and, more recently, globin X (in fish and amphibians [1]) and globin Y (specific to amphibians [2]).

Haemoglobin genes (alpha- and beta-globin) are of particular interest because of their critical role in oxygen transportation from the respiratory surfaces to the inner organs, and because of the dire effects of mutations in human globin genes that cause haemoglobinopathies [3]. The genes contained in the alpha (α)- and beta (β)-globin clusters are expressed at different stages of development and in different tissues. Together, gene products from both clusters form the functional tetrameric haemoglobin molecules needed to fulfil oxygen requirements.

The evolutionary history of α- and β-globin genes can be traced back to the common ancestors of fish, amphibians and amniotes (reptiles, birds and mammals), by comparing gene structure and composition of α- and β-globin clusters across vertebrates. In the amphibians *Xenopus laevis* and *X. tropicalis*, α- and β-globin genes are tightly juxtaposed as 5'-α-β-3' [2,4-6]. In the Antarctic notothenioid fish (*Notothenia coriiceps, N. angustata, Trematomus hansoni, T. pennellii*), there is also a single 5'-α-β-3' locus [7], although in pufferfish (*Fugu rubripes*) there are two globin clusters (one with α-globin genes and the other with both α- and β-globin genes), which are located on different chromosomes [8].

In amniotes, α- and β-globin clusters are located on different chromosomes. It was proposed that the ancestral α- and β-globin genes were located together in the common ancestor of amniotes, as they are in fish and amphibians, but became separated, either by chromosome fission or translocation between α- and β-genes, or by chromosome/genome or *in trans* duplication and gene loss [5].

Further duplications then occurred in amniote lineages. The ancestral α-globin gene is thought to have duplicated twice before the divergence of the bird-mammalian lineages, to produce progenitors of embryonic globin genes π/ζ, and adult $\alpha^D$ and $\alpha^A$, all of which are present in birds (for example, the chicken *Gallus gallus*) [9-11] and mammals [12,13]. The order and timing of these duplications is still debated, as is their origin: for instance, $\alpha^D$ may have evolved by duplication either of adult $\alpha^A$ (see [12]), or of an embryonic α-like gene [14]. After the avian and mammalian lineages diverged, there were further tandem duplications of the π/ζ and $\alpha^A$ lineages to produce more complex marsupial and eutherian ('placental') mammalian α-globin clusters, 5'-ζ-ψζ'-$\alpha^D$-ψα$^3$-α$^2$-α$^1$-θ-3' (see [12,15-18]). The timing of these duplication events is also uncertain, because we do not know whether these seven α-like globin genes all existed at the stem of the mammalian radiation.

As for many other gene families [19], comparisons of globin genes between distantly related mammals have provided unique insight into the evolution and function of the mammalian globins. Marsupials diverged from eutherian mammals about 148 million years ago (MYA), and mammalian Subclass Theria that contains these groups diverged from monotremes (Subclass Prototheria) about 166 MYA [20], so comparisons between these major mammal groups provide depth for evolutionary comparisons. Monotremes retain many anatomical and developmental features shared with birds and reptiles. Their small genome, too, and disjunct chromosome size classes are reminiscent of reptile genomes, and the 10 sex chromosomes in a karyotype of 52 chromosomes is unique among mammals [21-23]. Their importance for comparative studies is now increasingly recognised after the sequencing of the genome of a monotreme, *Ornithorhynchus anatinus* (platypus), to a depth of six to eight times by the Washington University Genome Centre, St Louis [24].

Indeed, studies of marsupial globins have clarified the timing of some of the duplications. The finding of single ε- (embryonic) and β-globin (adult) genes together in the marsupial β-globin cluster indicated that a two-gene cluster (ε-β) was present in the common therian ancestor [25-28]. Genes in the cluster were further duplicated to produce the ancestral eutherian β-globin cluster of 5'-ε-γ-η-δ-β-3' (see [29-32]), which then underwent further tandem duplication events. In contrast, the bird (*G. gallus*) β-like globin genes (ε-$\beta^H$-$\beta^A$-ρ) show very little homology to the mammalian β-like globin genes [33,34].

The discovery of a β-like globin gene (ω -globin) adjacent (3') to the α-globin cluster in marsupials led to a re-interpretation of globin evolution in birds and mammals [35,36]. Comparative sequence and phylogenetic analysis suggested that the ω-globin gene was more closely related to bird β-like globin genes than to other mammalian β-like globin genes. The specific function of the ω-globin gene is not yet known, but it is expressed just before birth and in the early stages of pouch young development [37]. In addition, the ω-globin product binds to α-like globin chains to form functional haemoglobin, so it is likely to be involved in oxygen transportation [35-37].

This finding of a remnant β-like globin gene (ω -globin) beside the α-globin cluster in marsupials [35,36] provided some support for the alternative hypothesis [5] that the α-

and β-globin clusters in birds and mammals arose by *in trans* duplication of a chromosomal region, rather than simply by separation of the ancestral α-β globin cluster by chromosome fission or translocation. Wheeler et al. [35,36] proposed that before the divergence of birds and mammals (>315 MYA), the chromosome region bearing the ancestral α-β clusters duplicated to form two clusters (α1-β1 and α2-β2) on different chromosomes, and their contents diverged independently in mammals and birds by silencing of some genes within each cluster (Figure 1). To account for the apparent orthology of the marsupial ω-globin gene and bird β-like globin genes, Wheeler et al. [35,36] suggested that the α1 and β2 were silenced in the eutherian lineage, but β2 was retained in marsupials as the ω-globin. In contrast, α2 and β1 were silenced in the bird lineage (Figure 1). On this hypothesis, then, both the α clusters and the β clusters of birds and mammals are paralogous (that is, evolved independently from ancient duplicates in an amniote ancestor) rather than orthologous (that is, diverged from the same ancestral cluster in an amniote ancestor).

This paralogy hypothesis (which rests on the rather weak orthology between the chicken β and marsupial ω), as well as the dates and types of other duplications, could be further tested by studying globin genes of monotreme mammals, and using comparative data to infer the ancestral globin gene arrangement of a mammal ancestor 166 MYA. The availability of platypus genomic sequences now provides an efficient way to discover all of the globin genes and regulatory signals, and to understand their function and evolution. Studies of globin genes in monotremes are also interesting because the specialized features and lifestyle of these unique mammals may have given rise to special adaptations of globin genes to fulfil unusual oxygen requirements. These features include the need for oxygen by diffusion through the egg membrane to the embryo after birth and the physiological response to hypoxic conditions during hibernation, burrowing and diving [38-40].

Little is known about monotreme α- and β-globin families. More than 30 years ago, studies of adult blood revealed a single adult α and β globin protein in the plat-

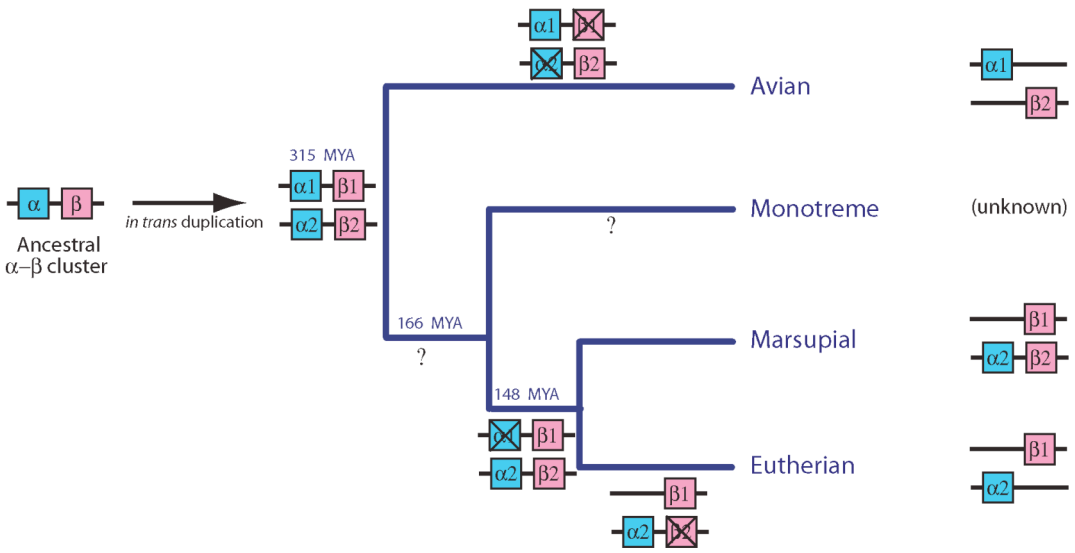



**Figure 1**
**Current proposed model for the evolution of α- and β-globin clusters from paralogous clusters in different lineages**. The unlinked α- and β-globin clusters in birds and mammals evolved from an ancient *in trans* duplication of the ancestral linked α-β cluster, followed by differential gene silencing (marked with X). This resulted in bird β-like globin genes (β2) orthologous to the marsupial ω-globin gene (β2 beside the α-globin cluster) but paralogous to mammalian β-like globin genes (β1). Adapted from Wheeler et al. [36].

ypus [41,42] and echidna (*Tachyglossus aculeatus* [43,44]). Lee et al. [28] later isolated an adult β-globin gene in the echidna that encoded a polypeptide identical to the previously isolated echidna β-globin [44]. To date, there is no evidence of any monotreme embryonic ζ- or ε-globin genes.

We used platypus genomic sequences from bacterial artificial chromosomes (BACs) to characterise the α- and β-globin gene families of the platypus and investigate their molecular evolution. In particular, we searched for embryonic and ω-globin genes and any novel globin genes that might fulfil the requirements for oxygen transport under hypoxic conditions. We investigated the genome context in order to infer the structure and origin of the ancestral α- and β-globin clusters at the stem of the mammalian radiation. Our results strongly support the hypothesis that the mammalian α- and β-globin clusters are orthologous to the avian α- and β-globin clusters, respectively, and that the β cluster evolved by transposition of a copy of the beta-like ω-globin gene in an amniote ancestor.

## Results

### *Identification of BAC clones containing the α- and β-globin clusters*

The draft sequence assembly of platypus [24] is readily available on the University of California Santa Cruz (UCSC) Genome Browser [45]. However, currently the assembly is incomplete for the α- and β-globin clusters, as individual globin genes appear on different contigs. There are also sequences of the platypus BAC clones available in NCBI GenBank that are not yet annotated and assembled, nor is part of the platypus genome assembly. Two of these are Oa_Bb-2L7 [GenBank:AC195438] and Oa_Bb-131M24 [AC203513], which were identified from the Encyclopaedia of DNA Elements Project to contain parts of the α-globin cluster (see Methods). The BAC clone Oa_Bb-484F22 [GenBank: AC192436] containing the β-globin cluster was obtained by screening a male platypus BAC library (Clemson University Genomic Institute, USA) and was subsequently fully sequenced and assembled by the Washington University Genome Sequencing Centre (St Louis, USA). These sequences were therefore used in this study to characterise the whole α- and β-globin clusters in the platypus.

Genes in these sequenced BAC clones were predicted by programs GENSCAN [46] and GenomeScan [47]. Many genes were predicted, which were then used for BLAST searches of nucleotide (BlastN) and amino acid (BlastP) databases to help identify them (data not shown). Phylogenetic analyses were also conducted for the platypus α- and β-like globin genes to further verify the identity of each gene (see below and also Figures 2, 3 and 4 below). With only one exception (platypus ε-globin, see below), the identities of all of the genes inferred by BLAST analyses were supported by phylogenetic analyses with high posterior probabilities and bootstrap support values.

### *Predictions and characterisation of genes in the platypus α-globin cluster*

One BAC (Oa_Bb-2L7) contained two embryonic α-like globin genes, and a second BAC (Oa_Bb-131M24) contained six α-like globin genes and a β-like globin gene (see Additional file 1). These two BACs were found to overlap by 10,066 base pairs (bp), resulting in a contig of 330,126 bp that contained the entire platypus α-globin cluster and flanking genes.

The 330,126 bp α-globin contig was found to contain six α-like globin genes, a β-like globin gene, and a gene that bore little similarity to α- and β-like globin genes but some similarity to cytoglobins (Figure 5A). These six α-like globin genes have a three-exon/two-intron structure and conserved donor/acceptor splice sites (GT/AG) typical of all vertebrate α-like globin genes. They are separated from each other by 2 to 6 kilobase pairs (kb). Full details of the exon/intron lengths, location of the putative poly-A addition site (AATAAA) and the lengths of the coding domains with the predicted encoded polypeptide for each predicted gene are given in Table 1. Figure 5B shows the predictions for some of the well-characterised protein-binding sites in the 5' promoter region (about 200 bp 5' to the cap site of each gene). These include CACCC [48], CAAT [49], TATA [50], GATA 1 [51], EKLF (Erythroid Krüppel-like Factor; [52]) and have been experimentally shown to control the stage- and tissue-specific expression of α- and β-like globin genes in other mammals [50,53-55].

Two genes at the 5' end of the α-globin cluster were both identified as ζ-like (referred to here as ζ and ζ') and predicted to encode polypeptides of 142 amino acids (aa), which are typical of known functional mammalian α-like globin genes. The amino acid sequence alignment of ζ and ζ' shows 95% identity. In the promoter region of both genes, CACCC and CAAT consensus boxes are conserved at similar positions, and in comparable order to that of human ζ and ζ' (Figure 5B).

Adjoining the two ζ-like globin genes, four other α-like globin genes were identified. One was an orthologue of bird and reptilian $\alpha^D$, and the other three were orthologues of adult α genes (here called $\alpha^3$, $\alpha^2$ and $\alpha^1$). The long and uninterrupted open reading frame (ORF) of $\alpha^D$ strongly suggests that it encodes a functional polypeptide of 141 aa, typical of known functional $\alpha^D$ globin genes. The platypus $\alpha^D$ globin gene contains introns of 1450 bp (intron 1) and 1610 bp (intron 2) that are very large com-

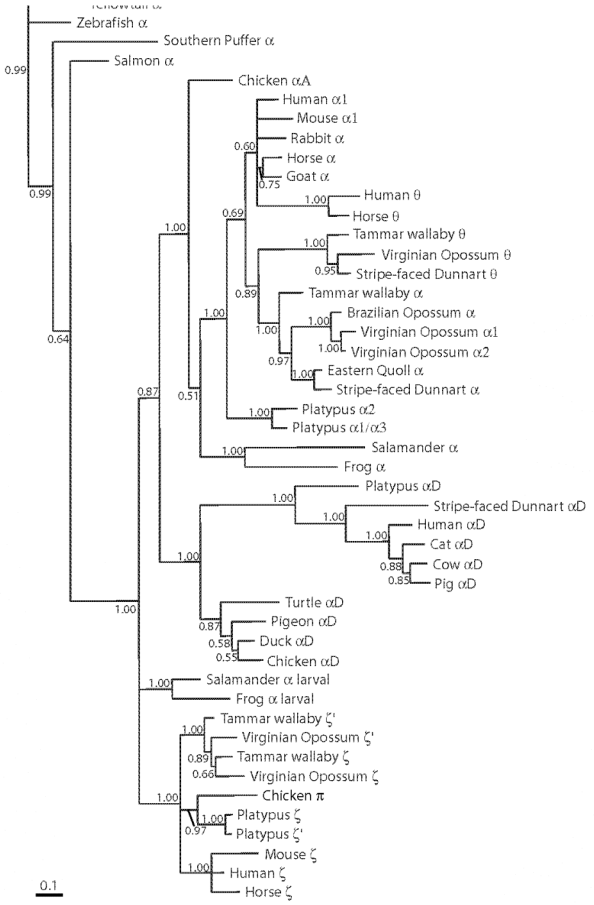



**Figure 2**

**Evolutionary relationships among vertebrate α-like globin genes using a 50% majority rule consensus phylogram from an analysis using Bayesian Inference**. The tree was constructed using mixed models of evolution for each codon position (see methods) and estimated base frequencies in an unlinked analysis using MrBayes (v. 3.1.2). Numbers adjacent to branches refer to % posterior probabilities. GenBank accession numbers for sequences are: Virginian Opossum (*Didelphis virginiana*) $\zeta^1$, $\zeta^2$, $\alpha^1$, $\alpha^2$, θ [AC139599.2, AC148752.1]; Stripe-faced Dunnart (*Sminthopsis macroura*) $\alpha^D$, $\alpha^2$, θ [AC146781]; Brazilian Opossum (*Monodelphis domestica*) α [TI# 453585430]; Tammar wallaby (*Macropus eugenii*) θ [AY459590], α [AY459589]; ζ [AY789121], ζ' [AY789122]; Horse (*Equus caballus*) θ (ψ α) [Y00284], $\alpha^1$ [M17902], ζ [X07051]; pig (*Sus scrofa*) $\alpha^D$ [AC145444]; cat (*Felis catus*) $\alpha^D$ [AC130194]; cow (*Bos taurus*) $\alpha^D$ [AC150547]; Goat (*Capra hircus*) α [J00043]; Human (*Homo sapiens*) α1 [V00491], θ [X06482], ζ [NM_005332]; *mu*/$\alpha^D$ chain [AY698022]; Mouse (*Mus muscularis*) $\alpha^1$ [NM_008218], ζ [X62302]; Rabbit (*Oryctolagus cuniculus*)α [X04751]; Eastern Quoll (*Dasyurus viverrinus*) α [M14567]; Chicken (*Gallus gallus*) $\alpha^A$, π, $\alpha^D$ [AF098919]; Duck (*Cairina moschata*) $\alpha^D$ [X01831]; Pigeon (*Columba livia*) $\alpha^D$ [AB001981]; Turtle (*Geochelone nigra*) $\alpha^D$ [SEG# AB1165195]; Zebrafish (*Danio rerio*) $\alpha^1$ [NM_131257]; Salamander (*Hynobius retardatus*) larval α [AB034756]; Salamander (*Pleurodeles waltlii*) α [M13365]; Frog (*Xenopus laevis*) α *1* [X14259], larval (tadpole) α T5 [X02798]; Yellowtail (*Seriola quinqueradiata*) $\alpha^A$ [AB034639]; Salmon (*Salmo salar*) α [X97289]; Southern Puffer (*Sphoeroides nephelus*) $\alpha^2$ [AY016023]; Platypus (*Ornithorhynchus anatinus*) ζ, ζ', $\alpha^D$, $\alpha^3$, $\alpha^2$, $\alpha^1$ [AC203513].

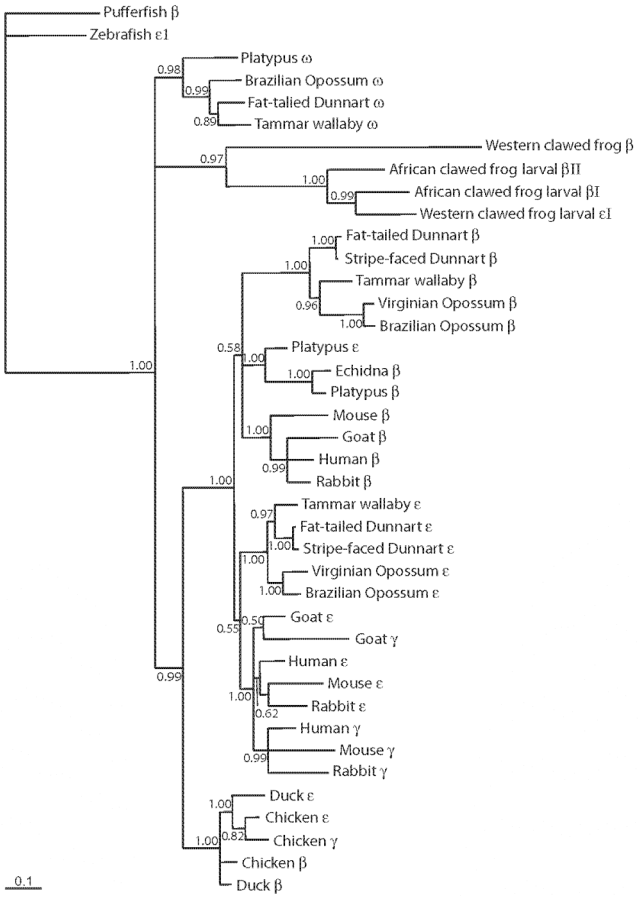



**Figure 3**

**Evolutionary relationships among vertebrate β-like globin genes using a 50% majority rule consensus phylogram from an analysis using Bayesian inference**. The tree was constructed using mixed models of evolution for each codon position (see methods) and estimated base frequencies in an unlinked analysis using MrBayes (v. 3.1.2). Numbers adjacent to branches refer to % posterior probabilities. GenBank accession numbers for sequences are: Fat-tailed Dunnart (*Sminthopsis crassicaudata*) β [Z69592], ε [Z48632], ω [AY014770]; Stripe-faced Dunnart (*S. macroura*) β, ε [AC148754]; Virginian Opossum (*Didelphis virginiana*) β [J03643], ε [J03642]; Brazilian Opossum (*Monodelphis domestica*) β [XM_001365299], ε [XM_001364448], ω [XM_001364828]; Tammar Wallaby (*Macropus eugenii*) β [AY450928], ε [AY450927], ω [AY014769]; African clawed frog (*Xenopus laevis*) larval β *I* [NM_001086273], larval βII [NM_001088028]; Western clawed frog (*X. tropicalis*) β [NM_203528], larval ε1 [NM_001016495]; Chicken (*Gallus gallus*) β ($\beta^A$) [NM_205489], ε [NM_001004390], γ ($\beta^A$) [NM_001031489]; Duck (*Cairina moschata*) β [J00926], ε [X15740]; Human (*Homo sapiens*) β [NM_000518], γ [BC130459], ε [NM_005330]; Mouse (*Mus musculus*) β (β1) [NM_008220], γ (β h0) [NW_001030869], ε ($\varepsilon^Y$) [M26897]; Goat (*Capra hirus*)*β* ($\beta^A$) [DQ350619], ε ($\varepsilon^I$) [X01912], γ [M15388]; Rabbit (*Oryctolagus cuniculus*) β, γ, ε [M18818]; Echidna (*Tachyglossus aculeatus*) β [L23800]; Pufferfish (*Fugu rubripes*) β [AY170464]; Zebrafish (*Danio rerio*) ε1 [NM_001103130]; Platypus β, ε [AC192436], ω [AC203513].

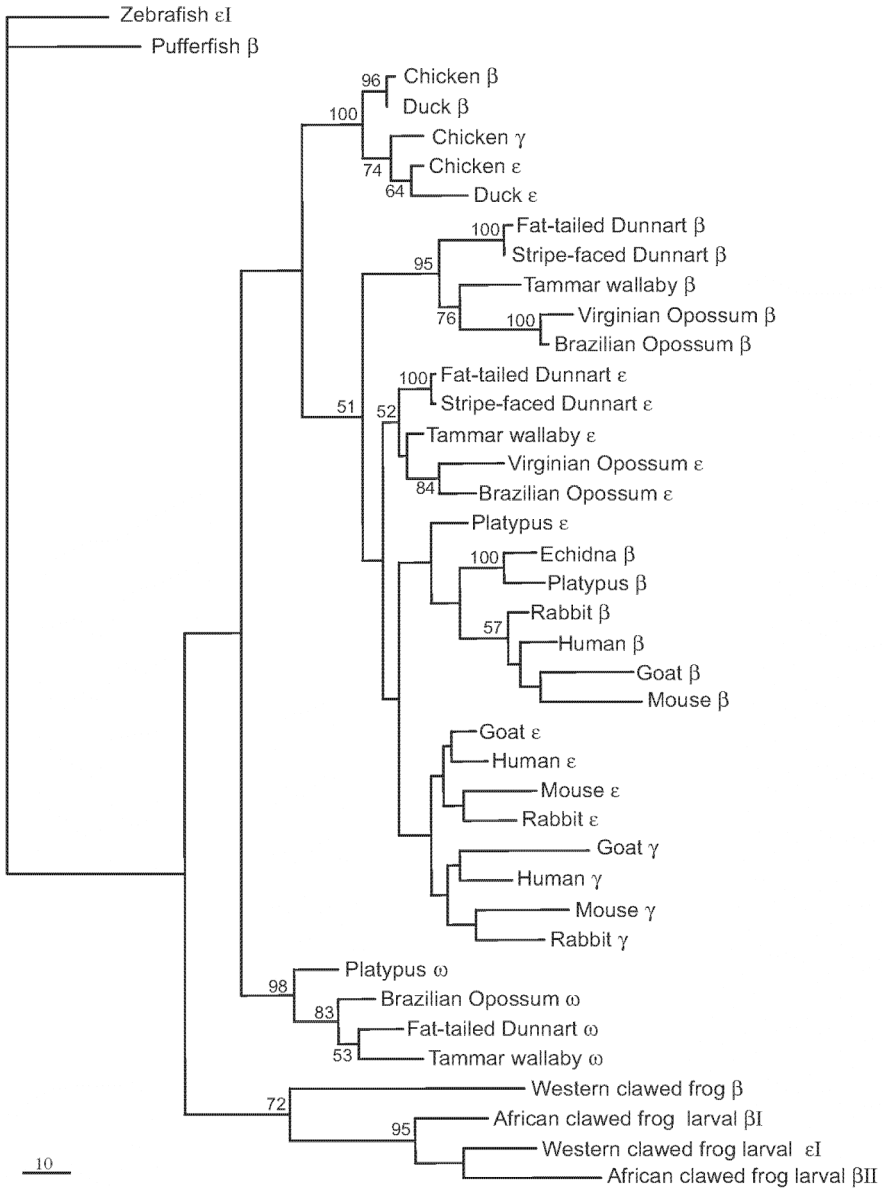



**Figure 4**
**Evolutionary relationships among vertebrate β-like globin genes analysed by maximum parsimony (MP) trees of length 926 (one of eight trees)**. Third position in codons were excluded in the MP analyses, which were conducted using a heuristic search in PAUP* v.4.0b10 [65]. The tree is rooted using pufferfish β-globin. Numbers adjacent to branches represent % bootstrap values (>50%) from MP heuristic analyses of 1000 pseudoreplicates. Accession numbers for sequences are given in the caption of Figure 3.

**Table 1: Gene-structure of the predicted platypus α- and β-like globin genes and *GBY***

| Length/ Genes | Exon 1 (bp) | Intron 1 (bp) | Exon 2 (bp) | Intron 2 (bp) | Exon 3 (bp) | Position of Poly-A | CDS (bp) | Poly-peptide (aa) |
|---|---|---|---|---|---|---|---|---|
| ζ | 95 | 337 | 205 | 114 | 129 | +119 | 429 | 142 |
| $\zeta^1$ | 95 | 336 | 205 | 102 | 129 | +133 | 429 | 142 |
| $\alpha^D$ | 92 | 1450 | 205 | 1610 | 129 | +77 | 426 | 141 |
| $\alpha^1/\alpha^3$ | 92 | 405 | 205 | 151 | 129 | +94 | 426 | 141 |
| $\alpha^2$ | 95 | 720 | 205 | 155 | 129 | +115 | 429 | 142 |
| ω | 92 | 256 | 223 | 111 | 129 | +69 | 444 | 147 |
| ***GBY*** | 98 | 3364 | 223 | 3053 | 144 | +141 | 465 | 154 |
| ε | 92 | 143 | 223 | 474 | 129 | +96 | 444 | 147 |
| β | 92 | 153 | 223 | 438 | 129 | +71 | 444 | 147 |

For each predicted gene, the length of the exons and introns, position of poly-A addition site (AATAAA) from the stop codon, and the length of their putative coding domain (CDS) and encoded polypeptide are shown. All genes contained consensus splice sites (GT/AG) in both introns.

pared with those of other α-like globins, which are usually less than 1000 bp.

Analyses of the platypus adult α-like globin genes reveal three adult ($\alpha^3$, $\alpha^2$ and $\alpha^1$) globin genes in the α-globin cluster. The sequence of $\alpha^3$ (the most 5' gene, adjacent to $\alpha^D$) was found to be almost identical to $\alpha^1$ (the most 3' gene) in their exon and intron regions, as well as in flanking regions of about 130 bp on both sides. The coding region was 100% identical, and just two sites in intron 1 were found to be different between the two genes. In order to confirm that identification of these two identical genes was not due to an error in the assembly of the original sequence data, the boundaries of the region containing the homology between $\alpha^1$ and $\alpha^3$ was further analysed by a BLAST search of the platypus whole-genome shotgun (WGS) database (data not shown). Two contigs were identified with homology to $\alpha^1$ and $\alpha^3$; these had identical sequences on one side of the boundary but different sequences on the other, confirming the presence of two separate genes. Further confirmation was obtained by performing a Southern blot on the α-globin-containing BACs, digested with an enzyme (*EcoRV*) that does not cut within the $\alpha^1$, $\alpha^2$and $\alpha^3$ (data not shown). Probing with $\alpha^1/\alpha^3$ revealed two bright bands, corresponding to $\alpha^1$ and $\alpha^3$, and one fainter band between them, corresponding to $\alpha^2$. Probing with $\alpha^2$ produced the same three bands, but in this case the middle one was brighter, corresponding to $\alpha^2$, and the outer bands were fainter, corresponding to $\alpha^1$ and $\alpha^3$. These analyses confirmed the existence of separate genes $\alpha^1$ and $\alpha^3$ in the platypus α-globin cluster. The $\alpha^2$ gene, located between $\alpha^1$ and $\alpha^3$, was distinct from both genes in the coding sequence (with 83% homology), in intron lengths (intron 1: 405 bp in $\alpha^1/\alpha^3$ and 720 bp in $\alpha^2$; intron 2: 151 bp in $\alpha^1/\alpha^3$ and 155 bp in $\alpha^2$) and in the promoter region (Figure 5B).

The amino acid sequence encoded by $\alpha^1$ and $\alpha^3$ was identical to the platypus adult α-chain previously identified by Whittaker and Thompson [41], implying that at least one of these genes is expressed in the adult platypus. The coding domain of $\alpha^1$ and $\alpha^3$ is shorter (426 bp) than that of $\alpha^2$ (429 bp), because it lacks the first three nucleotides of exon 1. The ORF of $\alpha^2$ gives a strong indication that it is translated into a functional polypeptide of 142 aa, typical of known functional mammalian α-like globin genes.

On the 3' side of the six α-like globin genes, a β-like globin gene was predicted, which was identified as the orthologue of the marsupial ω-globin gene. This platypus ω-globin gene has a typical three-exon/two-intron structure, conserved donor/acceptor splice sites, and encodes a polypeptide of 146 aa, typical of all vertebrate β-like globin genes (Table 1). The promoter region located 5' of the ω-globin initiation codon contains conserved sites for CAAT-EKLF-CACCC in an order identical to that of marsupial ω-globin gene.

Unexpectedly, GenomeScan predicted a gene based on the protein similarities with the α- and β-polypeptide chains, approximately 1.5 kb 3' of the ω-globin gene. Like other α- and β-globins, this gene also has a three-exon/two-intron structure and conserved donor/acceptor splice sites (Table 1). The lengths of its exons 1, 2, and 3 are 98, 223 and 144 bp, respectively, compared with 92, 223 and 129 bp in other β-like globin genes. However, it has much larger introns of 3364 bp (intron 1) and 3053 bp (intron 2). The long and uninterrupted ORF of this gene can be translated into a polypeptide of 154 aa, which is atypical of any known α- or β-like globin genes. A BLAST search of the amino acid sequence of this gene obtained the best hit with Globin Y (*gby*) of the amphibian *X. laevis* (identity score of 39%), and weaker identity scores with Cytoglobins (*cygb*) of other species, such as the fish *Danio rerio* (27%), *X. tropicalis* (26%), chicken (28%) and human

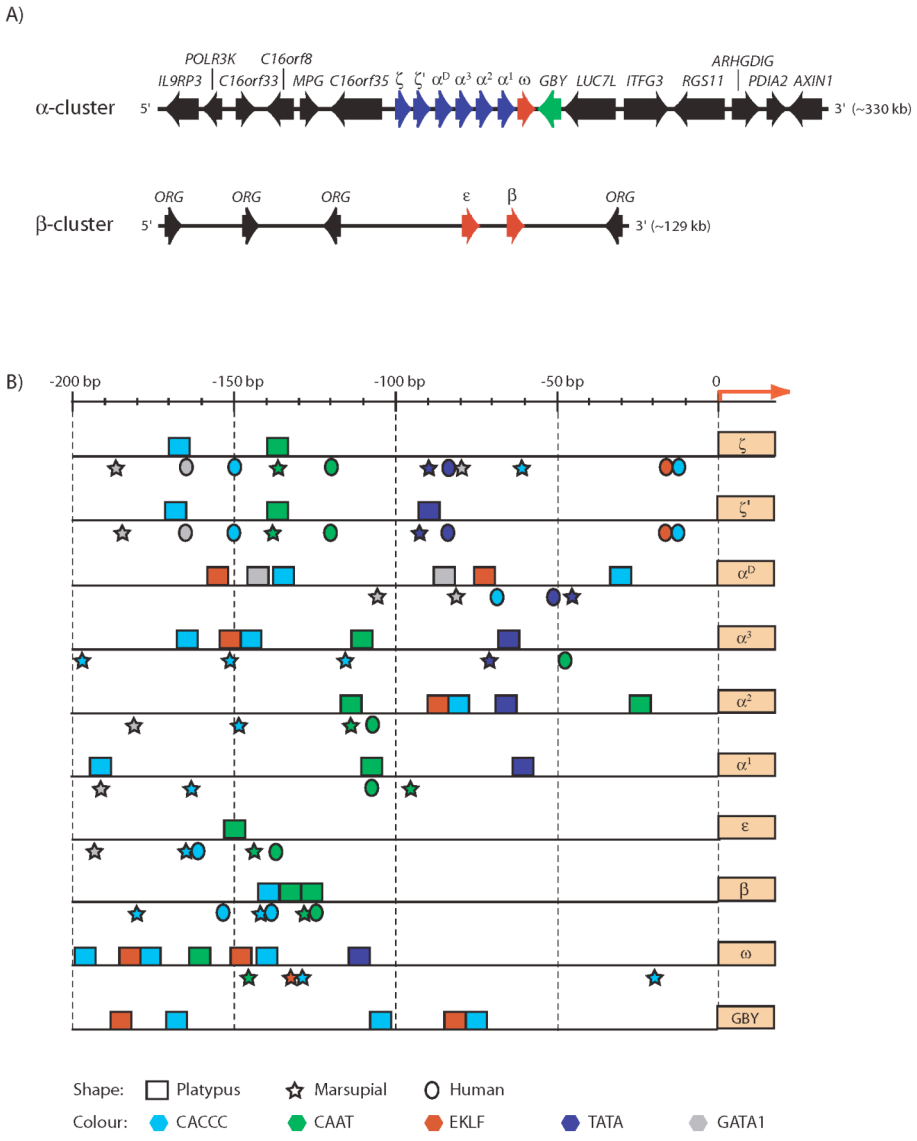



**Figure 5**
**Gene structure of the platypus α- and β-globin clusters and flanking loci, and comparisons of their promoter regions with other mammals**. (A) The platypus α-globin cluster contains six α-like globin genes (red), a β-like (ω) globin gene (blue) and a distantly related globin gene, *GBY* (green), which are flanked by *IL9RP3-POLR3K-C16orf33-C16orf8-MPG-C16orf35* on the 5' end and *LUC7L-ITFG3-RGS11-ARHGDIG-PDIA2-AXIN1* on the 3' end (black). The platypus β-globin cluster contains only two genes, ε and β (blue), which are flanked on both sides by *ORG* genes (black). (B) Relative positions of the putative transcription factor binding sites in the 200 bp promoter region located upstream of the predicted platypus, marsupial (*Didelphis virginiana* ζ and ψζ', and *Sminthopsis macroura* $\alpha^D$, $\psi\alpha^3$, $\alpha^2$, $\alpha^1$, ω, ε and β) and human α- and β-like globin genes. For the platypus *GBY* no data was available from other species, including *Xenopus tropicalis*, for comparisons.

(25%) at the protein level. We designated this gene '*GBY*' based on similarities with *X. laevis gby*, and its similar position adjoining the globin cluster [2]. The predicted polypeptide of platypus *GBY* (154 aa) was shorter than *X. laevis gby* (156 aa), and quite different from *X. laevis cygb* (179 aa), *D. rerio cygb1* (174 aa) and *cygb2* (179 aa), and human *CYGB* (190 aa). Using the Expressed Sequence Tag (EST) database, a BLAST search of the platypus *GBY* also obtained an identity score of 38% with *X. tropicalis gby* that was expressed in both tadpoles and adults, but produced no significant matches with any other mammalian genes. The present work was the first opportunity to analyse the promoter region of any *GBY* gene (Figure 5B).

### *Predictions and characterisation of genes in the platypus β-globin cluster*

In the platypus, only two β-like globin genes were predicted within the 129,521 bp BAC clone (Oa_Bb-484F22) by GENSCAN and GenomeScan (see Additional file 1). When the predicted amino acid sequences were subjected to BLAST search, the 5' gene had best hits with mammalian embryonic ε-globin genes. Although the phylogenetic analyses using Bayesian inference (BI; see below) indicated that this gene was more closely related to the platypus and echidna adult β-globin genes than to therian ε-globin genes, the position of this gene on the 5' end of the β-globin cluster and expression data (see below) supports its orthology with mammalian embryonic ε-globin genes, and is henceforth referred to as ε. The 3' gene encoded a protein identical to the previously identified platypus adult β-chain [42], and is henceforth referred to as β.

Both genes encode polypeptides of 146 aa, typical of known functional mammalian β-like globin genes. The promoter region of the platypus β has conserved sites of CACCC and CAAT in all three extant of mammals. However, the promoter region of the platypus ε appears to be quite different from other mammalian ε-globin genes and even from the platypus β (Figure 5B). The promoter of platypus ε contains only one predicted motif (CAAT), whereas the promoters of other mammalian ε, β and the platypus β contain many predicted motifs.

### *Expression studies of the platypus α- and β-like globin genes*

Transcription studies were performed to gain insight into the expression and function of all of the predicted platypus globin genes. Adult liver, kidney, spleen, testis, lung and brain were obtained for this project: no embryonic samples were available (or are ever likely to be available) for this vulnerable and iconic species. Observation of the expression of any of the predicted genes would constitute a good indication that the gene is transcriptionally active and functional.

Reverse-transcriptase polymerase chain reaction (RT-PCR) of all predicted platypus genes showed that they are all expressed in at least some of these adult platypus tissues (Figure 6). Platypus genes $\alpha^1/\alpha^3$, $\alpha^2$ and β, whose orthologues are usually expressed in the bone marrow of an adult human, were expressed in almost all platypus tissues tested, suggesting a broader expression of these genes in the monotreme lineage. Surprisingly, the genes ζ, ζ' and ε, whose therian orthologues are expressed only at embryonic stages of development, were expressed in adult spleen and testis, but not in the other tissues of adult platypus. This suggests that persistent expression of these genes in some adult tissues was selected for in the platypus, perhaps in response to its aquatic lifestyle and the hypoxic conditions of a confined burrow. Also, the expression pattern of platypus ε is similar to embryonic α-like ζ and ζ' but different from that of adult globin genes ($\alpha^1/\alpha^3$, $\alpha^2$ and β). The ω and $\alpha^D$ globin genes, whose functions are unknown, were also expressed mainly in the spleen. *GBY* was expressed in all adult platypus tissues, most strongly in testis.

### *Phylogenetic analyses*

Phylogenetic analyses of the α-like globin genes using BI and maximum parsimony (MP) produced several noteworthy results. The platypus adult α globin genes ($\alpha^1/\alpha^3$ and $\alpha^2$) grouped closely together to the exclusion of eutherian and marsupial α- and θ-globin genes for all analyses, although posterior probability (69%) and bootstrap support (66%) for this arrangement were relatively weak (Figure 2). This finding suggests that the duplication leading to the marsupial and eutherian θ-globin lineage occurred after the divergence of the monotreme and therian lineages. This is consistent with the absence of a θ-globin gene from the region between platypus $\alpha^1$- and ω-globin, its expected location based on its position in marsupial α-globin clusters [12,56].

Both platypus ζ-globin genes grouped closely together and formed a sister group relationship with chicken π, supported by a high posterior probability of 97% (Figure 2). A sister group relationship was also found in MP trees for analyses of the entire platypus coding region (bootstrap support <50%), and when third positions in the codon were excluded, was supported by 73% bootstrap pseudoreplicates (data not shown). This differs from the expectation that platypus ζ-globin genes would group with other mammalian ζ-globin genes to the exclusion of chicken π, suggesting that other factors (for example, purifying selection) operated to maintain a similar sequence in birds and monotremes.

There is still considerable uncertainty in the phylogenetic position of the $\alpha^D$-globin clade. It has recently been proposed that the $\alpha^D$ globin lineage resulted from duplica-

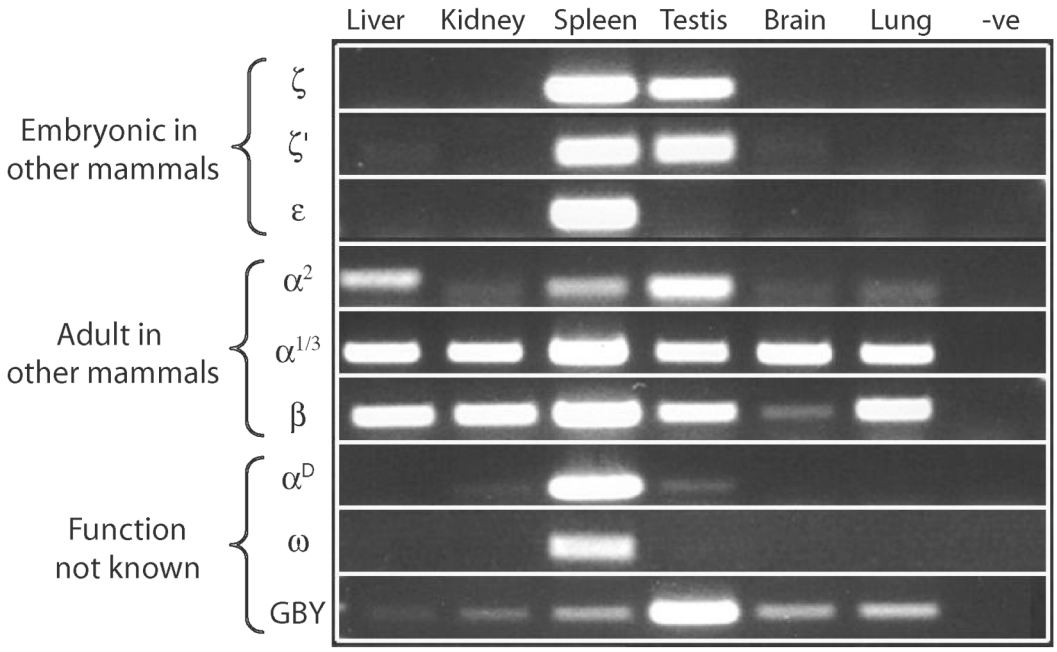



**Figure 6**
**Expression of all predicted α- and β-like globin genes including *GBY* in an adult platypus**. For each of the platypus predicted genes, expression was investigated by reverse transcriptase polymerase chain reaction in adult liver, kidney, spleen, testis, brain and lung. Primers for each gene were designed between two exons so that it would result in a product distinguishable from genomic contamination of cDNA. The negative control (last lane) contained no cDNA. All genes were expressed in one or more tissues, indicating that they are transcriptionally active and might be functional.

tion of the embryonic α-globin lineage, with phylogenetic analyses supporting a sister lineage relationship of these lineages to the exclusion of the adult α-globin lineage [14]. However, this arrangement was not supported in BI analyses of the data set used here, and the position of the $\alpha^D$ lineage was different in the different analyses. Analyses using BI (Figure 2) supported the sister lineage relationship of the $\alpha^D$ and adult α-globin lineages (as proposed by Cooper et al. [12]), with 87% posterior probability support. In contrast, all MP analyses supported the sister lineage status of $\alpha^D$ and embryonic α-globin genes, indicating an uncertainty in the phylogenetic position of the $\alpha^D$-globin clade.

Phylogenetic analyses of the β-globin genes provided results similar to recently reported phylogenetic analyses [35,36], with one notable exception. The BI analyses of coding sequence data (Figure 3) provided strong support (99% posterior probability) for the sister relationship of bird and mammalian β-like globin genes, contradicting previously published phylogenies of mammalian β-globin genes showing a sister relationship of marsupial ω-globin and bird β-like globin genes [35,36]. MP analyses (Figure 4), excluding third position in the codon, gave a similar tree arrangement, albeit with very low bootstrap support (<50%). In marked contrast to the BI analyses of DNA sequence data, BI protein analyses (data not shown) supported the sister relationship of bird β-like globin and mammal ω-globin lineages with a high posterior probability (99%).

Lastly, phylogenetic analyses using BI indicated that the platypus ε gene was more closely related to the platypus and echidna adult β-globin genes than to therian ε-globin genes, suggesting it may not be orthologous to marsupial and eutherian ε-globin (Figure 3). BI analyses of β-globin protein data and MP analyses of the coding sequence data, with third codon positions excluded, grouped the gene as

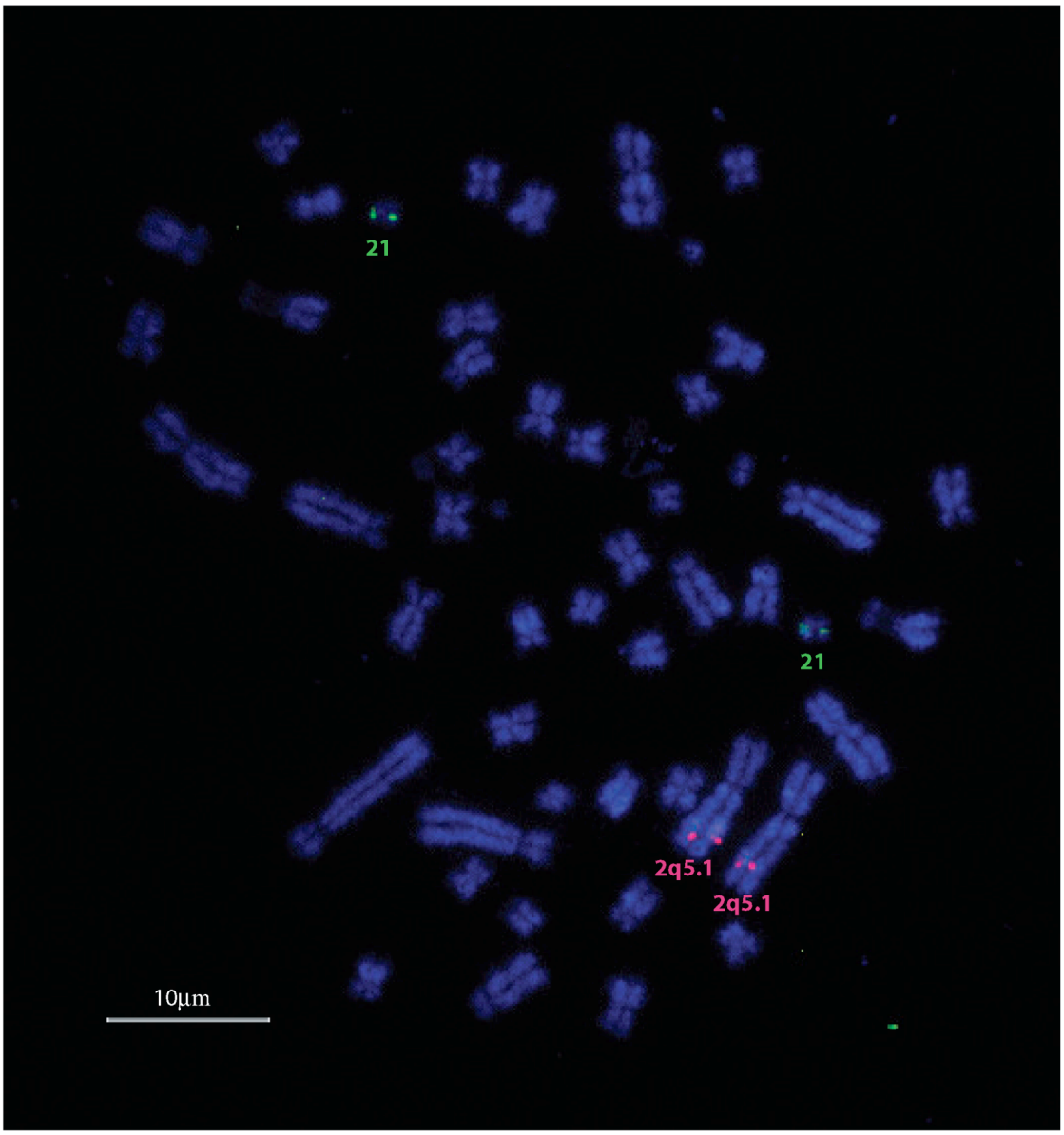



**Figure 7**
**Chromosomal location of the platypus α- and β-globin clusters**. Two-colour fluorescence in situ hybridisation showing the location of the α-globin cluster on chromosome 21 (green) and the β-globin cluster on chromosome 2q5.1 (red). The chromosomes are counterstained with DAPI (blue).

an ancestral lineage to eutherian and monotreme adult β-globin genes (see Figure 4). This ancestral position suggests that the lineage evolved following duplication of an ancestral β-globin gene prior to the divergence of monotremes and therians.

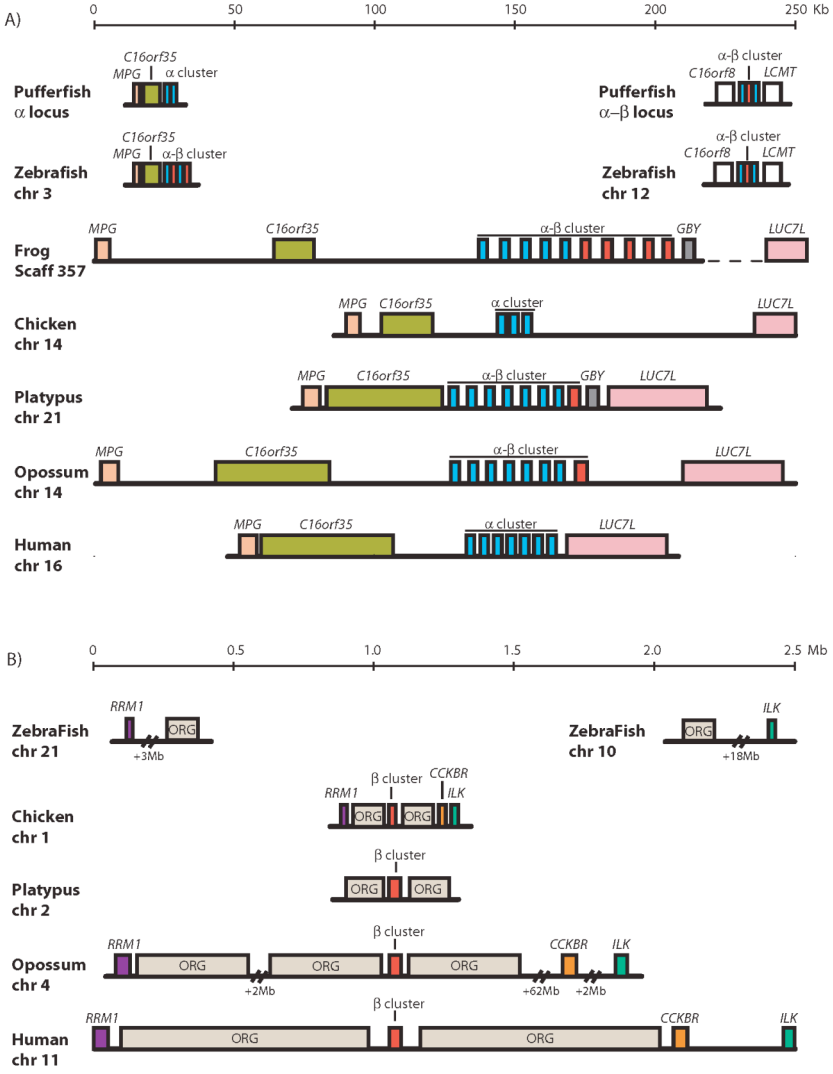



**Figure 8**

**Loci flanking vertebrate α-globin (A) and β-globin (B) clusters**. The relative locations of flanking loci (A) *MPG*, *C16orf35*, *LUC7L* and *GBY* and (B) *RRM1*, *CCKBR*, *ILK* and *ORG* genes were searched for beside the α-β globin cluster in zebrafish (*Danio rerio*) and frog (*Xenopus tropicalis*), and beside the separate α-globin and β-globin clusters in chicken (*Gallus gallus*), opossum (*Monodelphis domestica*) and human (*Homo sapiens)* from Ensembl [57]. The pufferfish (*Fugu rubripes*) flanking loci shown here were adapted from Gillemans et al. [8]. For the platypus, the α-globin flanking loci were characterised in this study, and *ORG* genes surrounding the platypus β-globin cluster were discovered: however, the BAC clone (484F22) was too small to cover the region containing the loci *RRM1*, *CCKBR* and *ILK*. In *X. tropicalis LUC7L* was found on another scaffold (466 from Ensembl) but sequence analyses by Fuchs et al. [2] suggested that *LUC7L* resides 3' to the frog α-β-*GBY* cluster. The flanking loci as well as the α- and β-globin clusters are differentiated by colour.

#### *Location of the α- and β-globin clusters in the platypus*

The location of the verified BAC clones containing the α- (Oa_Bb-2L7) and β-globin (Oa_Bb-484F22) clusters in the platypus was determined by fluorescence *in situ* hybridisation (FISH) (Figure 7). The β-globin cluster localised to one of the largest autosomes, giving unambiguous signals on the long arm of chromosome 2 (2q5.1). The α-globin cluster localised to the smallest autosome, 21, whose two arms are not distinguishable by size or DAPI banding pattern [21]. This is the first gene that has been localised on the platypus chromosome 21.

#### *Loci flanking the α- and β-globin clusters in the platypus and other vertebrates*

To explore the genome context of the α- and β-globin clusters in the platypus and other vertebrates, the platypus BAC sequences and the genomes of other sequenced species were searched for loci residing beside the α- and β-globin clusters.

As well as globin genes, GENSCAN predicted within the platypus α-globin 330,126 bp contig many genes that flank the platypus α-globin cluster (Figure 5A), which were identified by BLAST analyses. These include *IL9RP3-POLR3K-C16orf33-C16orf8-MPG-C16orf35* upstream (5') of the α-globin cluster, and, *LUC7L-ITFG3-RGS11-ARH-GDIG-PDIA2-AXIN1* downstream (3') of the α-globin cluster (Figure 5A).

To compare the α-globin flanking loci of the platypus and other vertebrates, the genes closest to the α-globin cluster, *MPG*, *C16orf35* and *LUC7L* were searched for in the human, opossum (*Monodelphis domestica*), chicken, frog (*X. tropicalis*) and zebrafish (*D. rerio*) genomes that were accessible from Ensembl [57]. Figure 8A shows that the locations of *MPG*, *C16orf35* and *LUC7L* are conserved adjacent to the α-globin cluster of birds and mammals, and in the same position adjacent to the α-β cluster of amphibians, and all but *LUC7L* were also present in fish. These results are consistent with the previous analyses of Flint et al. [58] and Hughes et al. [59]. Thus the flanking loci analyses reveal that the genome context of the platypus α-globin cluster is the same as the α-globin clusters in therian mammals and birds, and this is the same as for the α-β cluster of fish and frogs.

GENSCAN also predicted numerous genes other than globin genes in the platypus β-globin BAC (484F22). These were identified by a BLAST search as members of the olfactory receptor gene (*ORG*) family that are responsible for odour detection. Three conserved *ORG* members were identified at the 5' end of the platypus β-globin cluster and one conserved *ORG* member at the 3' end (Figure 5A).

To compare β-globin flanking loci, *ORG* genes, as well as other genes that are closest to the β-globin cluster in other species, *RRM1*, *CCKBR* and *ILK* were searched for in the human, opossum, chicken and zebrafish genomes that were accessible from Ensembl [57]. Data from frog (*X. tropicalis*) was not useful since all of these loci lie on different contigs or scaffolds due to assembly problems. The locations of multiple *ORG* genes, *RRM1*, *CCKBR* and *ILK* were found to be conserved adjacent to β-globin cluster of birds and mammals [60,61], but not for the α-β cluster of fish and frogs, nor beside the second α-β cluster of zebrafish and pufferfish (Figure 8B). Thus the genome context of the platypus β-globin cluster is the same as in therian mammals and birds, but this is different from the α-β cluster of fish and frogs.

## Discussion

The phylogenetic position of monotremes makes comparisons with platypus of special value for exploring the organization, function and evolution of mammalian genes and genomes. The availability of platypus genome sequence data now makes many such studies possible, and have been used here to characterise the platypus α- and β-globin gene clusters and explore their evolutionary history.

#### *The platypus α-globin gene cluster*

The platypus α-globin cluster contains at least eight genes within more than 40 kb, including six α-like globin genes (including the identical $\alpha^1$ and $\alpha^3$), one β-like globin gene (ω-globin) and a gene belonging to another member of the globin super-family (*GBY*) arranged in the order 5'-ζ-ζ'-$\alpha^D$-$\alpha^3$-$\alpha^2$-$\alpha^1$-ω-*GBY*-3' (Figure 5A). The cluster maps to chromosome 21, the smallest autosome in platypus. All eight genes are likely to be functional since their expression was detected in tissues of an adult platypus.

The platypus α-globin cluster is almost identical to the arrangement of α-like globin genes in the ancestral therian cluster reported by Cooper et al. [12]. The one exception is the absence of a θ-globin gene from the platypus cluster. Phylogenetic analyses support the basal position of the monotreme adult α-globin lineage relative to marsupial and eutherian α- and θ-globin lineages, implying that the duplication of an adult α-globin to produce θ-globin occurred in the therian lineage after its divergence from the monotreme lineage (Figure 9B). However, although the numbers and arrangements of genes is so similar in platypus and therians, the presence of three adult α-globin genes and two embryonic ζ-globin genes in their common ancestor was not supported by phylogenetic analyses, which showed independent groupings of the three adult and embryonic genes within each separate mammalian lineage (Figure 2 and see Cooper et al. [12]). This result can be interpreted literally as resulting from

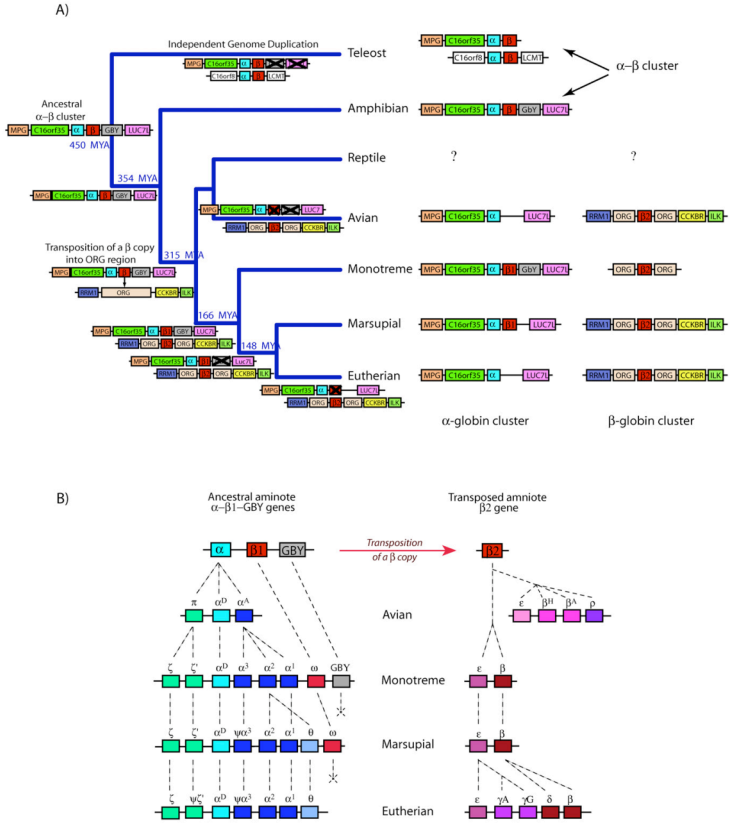



**Figure 9**

**Proposed model for the evolution of the α- and β-globin clusters in vertebrate lineages**. (A) A region containing *MPG-C16orf35*-α-β-*GBY-LUC7L* represented the ancient α-β globin cluster of jawed vertebrates (>450 MYA), which is seen in the amphibian lineage. This region further duplicated and underwent some gene silencing in teleost fish. In an amniote ancestor of reptiles, birds and mammals (>315 MYA), a copy of an ancestral β-globin gene from this region was inserted into a different chromosome within a region replete with multiple copies of *ORG* genes. The original amniote β-globin gene survives as the ω-globin gene (β1) in the α-globin cluster of marsupials and monotremes, whereas the transposed β-globin gene (β2) duplicated several times to form different clusters in the different lineages. (B) Tandem duplications of the ancestral amniote α-globin gene produced a three-gene ($\pi$-$\alpha^D$-$\alpha^A$) cluster in the avian lineage. In the mammalian lineage, further duplications gave rise to a six-gene ($\zeta$-$\zeta'$-$\alpha^D$-$\alpha^3$-$\alpha^2$-$\alpha^1$) cluster with ongoing gene conversion events homogenising the embryonic and adult genes. In monotremes, the ancestral ω (β1) and *GBY* are retained. After the divergence of monotreme and therian mammals, there was an additional duplication of $\alpha^2$ to form θ, giving rise to the seven-gene cluster ($\zeta$-$\zeta'$-$\alpha^D$-$\alpha^3$-$\alpha^2$-$\alpha^1$-θ) in marsupials and eutherians. Marsupials also retain the ancestral ω but may have lost *GBY* gene; eutherians retain no identifiable remnant of either gene. Furthermore, the ancestral transposed β2-globin gene duplicated independently in birds and mammals. Before the mammalian radiation, we propose that the ancestral β2 gene duplicated to form a two-gene β-globin cluster (ε-β) as seen in monotremes and marsupials, except that ongoing gene conversion events homogenised platypus ε to group with monotreme β genes. After the divergence of marsupial and eutherian mammals, there were further tandem duplications of these two genes to produce complex β-globin cluster (ε-γ-η-δ-β) in eutherians.

independent duplications in each mammalian lineage to produce three adult and two embryonic genes in each. However, this seems unlikely to explain the convergence in gene number of the α-globin cluster in these distantly related mammalian lineages. We suggest that a more parsimonious explanation is that the common ancestor of monotremes and therians contained three adult α-globin genes and two ζ-globin genes, which were homogenised by ongoing gene conversion events, leading to the gene tree that does not match the duplication history of the individual genes. The close similarity of the platypus $\alpha^3$ and $\alpha^1$ loci suggests a very recent gene conversion event that homogenised their sequences. Therefore, we propose that the platypus α-globin cluster of eight genes (ζ-ζ'-$\alpha^D$-$\alpha^3$-$\alpha^2$-$\alpha^1$-ω-*GBY*) represents the ancestral mammalian α-globin cluster arrangement (Figure 9B), in which all genes were transcriptionally active.

Importantly, the platypus α-globin cluster contains a copy of the β-like ω-globin gene, also found in the marsupial α-globin cluster, but absent in humans, supporting the hypothesis that ω-globin was present in the common ancestor of all mammals. Phylogenetic analyses also confirm the ancient ancestry of the ω-globin gene, as concluded by Wheeler et al. [35,36]. Among adult platypus tissues this gene was expressed only in the spleen. In marsupials, expression of the ω-globin gene was detected just prior to birth and during early pouch young development [37], although the site of expression was not studied, and there was no evidence of adult expression in blood cells.

#### *Discovery of a mammalian GBY globin gene adjoining the α-globin cluster*

We discovered a globin gene *GBY* in the platypus that is adjacent (3') to ω in the α-globin cluster. It has a typical three-exon/two-intron structure like other α/β-globin genes, contains an ORF encoding a polypeptide chain of 154 aa, and is expressed in almost all adult tissues, most strongly in testis. The amino acid sequence is unrelated to any of the other globin genes in the cluster, so it is unlikely to be derived by duplication of α- or ω-globin within the monotreme lineage. Rather, it shows sequence similarity to *gby* of *X. tropicalis* and *X. laevis*, a gene thought to be related to cytoglobins [2].

Little is known of the function of amphibian *gby*, or its relationship with other globins. Fuchs et al. [2] reported that amphibian *gby* encodes a *bona fide* globin of 156 aa, having all of the sequence features of a functional respiratory protein. *gby* was expressed in all adult tissues tested in *X. laevis*, most strongly in ovary, kidney and eye, and was present in 20 expressed sequence tag clones from different stages of *X. laevis* and *X. tropicalis* embryonic and adult development [2], suggesting that it is expressed in embryonic as well as adult stages. Phylogenetic analysis of all vertebrate globins [2] showed that the *gby* lineage diverged at the base of two separate clades, one comprising all vertebrate cytoglobins, myoglobins, agnathan globins and bird globin E, and the other comprising the haemoglobin α- and β-chains.

The position of platypus *GBY* adjacent to the α-globin cluster and flanked by *LUC7L* mirrors its position in *X. tropicalis* between the main α-β cluster and *LUC7L* [2]. Another common feature of both was strong expression in gonads (ovary in *X. laevis* [2] and testis in platypus), so *GBY* has sex-related expression in both lineages. Thus *GBY* is not specific to amphibians, as was thought, but was a component of the cluster in an ancient tetrapod, and has been lost, or has diverged beyond recognition, in birds and therian mammals.

#### *The platypus β-globin gene cluster*

Characterisation of the platypus β-globin cluster revealed two β-like globin genes over about 13.2 kb that are arranged in the same order as marsupials, 5'-ε-β-3' (Figure 5A). This cluster is located on platypus chromosome 2q5.1. Both genes appear to be transcriptionally active and are likely to be functional.

At the time of revising this paper, an independent paper on monotreme β-like globin genes was published by Opazo et al. [62] in which they reported the presence of ω, $\varepsilon^P$ and $\beta^P$ in the platypus. Largely on the basis of phylogenetic analyses of flanking and coding sequence data, they proposed that platypus $\varepsilon^P$ and $\beta^P$ were not 1:1 orthologues of therian ε and β, respectively, and arose by independent duplication of an ancestral β-globin gene in the monotreme lineage, with a separate duplication event, just prior to the divergence of therians, producing the progenitors of ε and β of therians. This hypothesis was strongly supported by our BI phylogenetic (Figure 3) analyses, but not by MP analyses of coding sequence data, with third codon sites excluded (Figure 4), or BI analyses of protein sequence data (not shown). These contradictory analyses highlight the difficulty in resolving deep relationships among globin genes, particularly when the time periods between duplication and speciation events are relatively small, the phylogenetic signal at third codon positions is potentially saturated, and non-synonymous sites may be subjected to purifying or positive selection. Despite a very high posterior probability (100%) for the grouping of platypus ε with monotreme β, this value is a Bayesian probability and depends on the model adequately representing the evolution of the gene. Furthermore, although it was reported [62] that the 5' flanking sequences of platypus ε and β were similar, we found no evidence for similarity of the promoter signals of these two genes (Figure 5B).

We consider that a more parsimonious explanation is that the platypus ε is orthologous to the marsupial and eutherian embryonic β-like globin lineages (ε and γ), and arose by duplication of an ancestral β-globin gene prior to the mammalian radiation (166 MYA; Figure 9B). The sequence of platypus ε may have been homogenised by some gene conversion events, leading it to group with other monotreme adult β-like globin genes. In addition to the MP analyses reported above, this explanation is further supported by the conserved position of ε to the 5' side of the adult β-globin gene in the platypus cluster, which is similar to that found in other therian β-globin gene clusters [26]; see also [29]). Amino acid sequence analyses (BlastP) also provided additional support for the orthology of platypus ε to other mammalian ε-globin genes. Although we were unable to examine the expression of the genes in embryonic tissues, it was found that the expression profile of the platypus ε was similar to the embryonic α-like globins ζ and ζ' of the platypus, but not to the adult β-globin gene, supporting its potential role as an embryonic β-like globin gene.

#### *The ω-globin gene and the evolution of the β-globin cluster*

The discovery of the marsupial ω-globin gene in the α-globin cluster [35,36] was critical in re-interpreting the relationships of the α- and β-globin clusters in amniotes (reptiles, birds and mammals) to favour the hypothesis that these clusters in birds and mammals are paralogous, having diverged independently from different ancestral copies of the vertebrate α-β-globin locus [63].

Our observation of an ω-globin gene in the α-globin cluster in the platypus, as well as in the marsupials, confirms that the ancestral mammal α-globin cluster contained a β-like globin gene that was lost in eutherians, as proposed by Wheeler et al. [35,36]. However, the position of monotreme and marsupial ω in the phylogeny (Figure 3) is more consistent with the original hypothesis [5] that mammal and bird β-globin are orthologous, having descended from the same β-globin progenitor in an amniote ancestor, and this is strongly supported by flanking sequence data (see below). Our data support the proposition that the ω -globin gene represents an ancient β-like globin gene lineage that is ancestral to a group containing both mammalian and bird β-globins with a high posterior probability (99%). This arrangement, however, was not supported by analyses of amino acid sequence data, indicating that there is uncertainty in the phylogenetic position of ω-globin relative to bird β-globins, or that convergent evolution of bird β-globin genes and ω-globin resulted in their similarity at the protein level. To further resolve the key question of whether bird and mammal β-globin gene clusters are orthologous we carried out comparative analyses of flanking loci of the α- and β-globin clusters.

#### *Genome context of vertebrate α- and β-globin clusters*

We found that the platypus α-globin cluster is flanked by *MPG*, *C16orf35*, *GBY* and *LUC7L*, and that the same genes (except *GBY*) flank the α-globin cluster in mammals and birds [58,59]. The same genes flank the α-β cluster of frog, and even zebrafish and the α-cluster of pufferfish [8] (except *GBY* and *LUC7L*), implying that a very ancient region containing these genes (5'-*MPG*-*C16orf35*-α-β-*GBY*-*LUC7L*-3'), or perhaps an even larger region, was present in their common ancestor and has been conserved since the evolution of jawed vertebrates more than 450 MYA.

In contrast, the amniote β-globin clusters reside in a very different genome, sharing none of the flanking loci with the mammal and bird α-globin clusters, or the α-β cluster of frogs and fish. In platypus, as well as in therian mammals [60,61], the β-globin clusters are flanked by numerous *ORG* genes on both sides. In birds, also, the β-globin cluster is embedded in *ORG* genes [60]. Even the outside loci *RRM1*, *CCKBR* and *ILK* lie in the same orientation with respect to the bird and mammalian β-globin clusters [60], suggesting that the 5'-*RRM1*-*ORG*-*β* (cluster)-*ORG*-*CCKBR*-*ILK*-3' arrangement has been conserved since before the divergence of birds and mammals, more than 315 MYA. Therefore, the bird β-globin cluster is orthologous to the β-globin clusters of mammals.

## Conclusion

#### *New model for the evolution of α- and β-globin clusters in amniotes*

This analysis of flanking loci, in addition to the phylogenetic analyses reported above, refutes the prevailing hypothesis that mammal and bird α- and β-globin clusters evolved from different (paralogous) copies of an ancestral α-β-globin region containing *MPG*-*C16orf35*-*α* (cluster)-β (cluster)-*GBY*-*LUC7L*. Rather, the context of β-globin clusters within olfactory receptor genes in birds as well as mammals suggests that a copy of a β-globin locus was moved into a region replete with *ORG* genes before the divergence of birds and mammals 315 MYA. The precise mechanism for this translocation is unknown, but is likely to be either by transposition of a tandem duplicate of an ancestral β-globin gene, or retrotransposition of an intron-containing primary transcript. Phylogenetic analyses suggest that this ancestral β-globin gene within the α-globin cluster is represented by the platypus and marsupial ω-globin gene. The transposed β-globin gene then independently duplicated several times within the avian and mammalian lineages to form the different clusters of differentially expressed β-globin genes. Full details of this new model are given in Figure 9A and 9B.

This hypothesis could be further tested by investigating the gene organization of the α- and β-globin clusters in

reptiles such as lizards and snakes, which form a sister group to birds. Our hypothesis predicts that reptiles should possess a *MPG-C16orf35*-α (cluster)-β (cluster)-*GBY-LUC7L* cluster, and an unlinked *RRM1-ORG-β* (cluster)-*ORG-CCBKR-ILK* cluster like birds and mammals. The full genome sequence of the first reptilian species,*Anolis carolinensis*, will provide an opportunity to test this hypothesis.

## Methods

### *Isolation and purification of probes to screen for platypus β-globins*

At the start of this project there were no trace sequences available for any globin genes in the platypus trace archive. We therefore designed probes to screen the platypus male Oa_Bb BAC library (Clemson University Genomic Institute, USA). The platypus β-globin-specific primers OaBGF (5'-tggacccagaggttctttgac-3') and OaBGR (5'-tgcaattcactcagcttggag-3') were designed from the reference tammar β-globin sequence [GenBank: AY450928] using Primer3 [64]. Amplification by PCR was performed in a final volume of 25 μl, with 40 ng genomic DNA, 1× Buffer (Roche, Australia), 0.2 mM dNTPs, 0.05 U Taq (Roche, Australia) and 1 μM each of forward and reverse primers. PCR cycling conditions were: 94°C for 2 minutes, then 35 cycles of 94°C for 30 seconds, 50 to 60°C for 30 seconds, 72°C for 1 minute, followed by 72°C for 10 minutes. The PCR products were sub-cloned according to the TOPO TA cloning® Kit Protocol (Invitrogen, Australia) and the resulting plasmids were purified according to the centrifugation protocol of Wizard® Plus SV Minipreps DNA Purification System (Promega, Australia). The purified plasmids were confirmed to contain PCR products of a partial platypus β-globin gene (167 bp) by sequencing at the Australian Genome Research Facility (AGRF, Brisbane, Australia) using M13 forward (5'-gtaaaacgacggccag-3') and M13 reverse (5'-caggaaacagctatgac-3') primers. Once confirmed, they were used as probes to screen the platypus BAC library.

### *Screening the platypus BAC library for β-like globin genes*

The platypus BAC library filters were pre-hybridised at 65°C with Church Buffer (1 mM EDTA, 0.5 M phosphate buffer, 7% (w/v) SDS) including 1% BSA for 4 hours. The platypus β-globin probes (25 ng) were labelled with $^{32}$P-dATP using the Megaprime DNA labelling System (GE Healthcare, Australia) following the manufacturer's instructions. The probes were allowed to hybridise to the filters at stringent conditions (65°C with the above buffer) for 24 hours and then washed twice for 15 minutes each in 2 × SSC/0.1%SDS and 1 × SSC/0.1%SDS. Autoradiography was carried out for 14 days at -80°C with an intensifying cassette.

### *Identification of platypus BAC clones containing α-like globin genes*

Unlike β, BACs were not screened for α-like globin genes. Instead they were identified directly from the Encyclopaedia Of DNA Elements Project [65], in which the α-globin cluster is one of the targeted regions [12,66]. Two platypus BAC clones (Oa_Bb-2L7 and Oa_Bb-131M24), which were sequenced but not yet annotated, were identified by computational analysis (below) to contain parts of the α-globin cluster and a ω-globin gene.

### *Isolation and purification of DNA from BAC clones*

DNA from the identified BAC clones (including those that were screened) was extracted using Wizard® Plus SV Minipreps DNA Purification System (Promega, Australia). The purified BAC clones were then subjected to Dot or Southern Blot to confirm the presence of α- or β-globin genes respectively.

### *Confirmation of BACs containing globin genes*

Dot blot methods were used to verify the presence of the α-like globin genes. In a plate containing Luria broth agar with chloramphenicol, a Hybond N$^+$ (GE Healthcare, Australia) filter was placed and multiple 1 μl of liquid culture BAC clones were spotted onto the filter. The plate was incubated at 37°C overnight and then the filter was soaked in Denaturation Solution (0.5 M NaOH and 1.5 M NaCl) for 5 minutes, followed by soaking twice in Neutralisation Solution (0.5 M Tris-Cl pH 7.4 and 1.5 M NaOH) for 5 minutes each. The filter was then rinsed in 2 × SSC, soaked in 0.4 M NaOH for 20 minutes and washed with 6 × SSC to remove all cellular debris. The filters were then screened with the platypus α-globin probes using the standard library screening procedure (above).

Southern blotting was used to verify the presence of the β-like globin genes. In a 40 μl reaction, 20 to 40 ng BAC DNA was digested with 10 U of restriction enzyme,*HIND III* (Roche, Australia). The reaction was incubated at 37°C for at least 4 hours and separated by electrophoresis on a 0.8% agarose gel overnight at 40 V. The DNA fragments were transferred onto a Hybond N$^+$ (GE Healthcare, Australia) nylon filter overnight by capillary action following the manufacturer's instructions, and cross-linked in 0.4 M NaOH for 20 minutes. These filters were then screened with the platypus β-globin probes using the standard library screening procedure (above).

### *Fluorescence* in situ *hybridisation (FISH)*

Male platypus metaphase spreads were prepared and *in situ* mapping was performed using two-colour FISH as described previously by McMillan et al. [21]. The verified BACs containing the α-like globin genes (ζ and ζ': Oa_Bb-2L7) and β-like globin genes (ε and β: Oa_Bb-484F22) were labelled with different fluorochromes and then

hybridised to the chromosomes. The signals were detected by fluorescent microscopy, where at least twenty metaphase images were captured and analysed.

#### *Sequence data of the platypus BAC clones containing the α- and β-globin clusters*

Information about the platypus BAC clones containing the α-like globin genes along with the ω-globin gene was obtained directly from the ENCODE Project [66]. Their sequence information was obtained from GenBank; accession numbers: AC195438 (Oa_Bb-2L7) and AC203513 (Oa_Bb-131M24).

The BAC clone containing the β-like globin genes that were found from the library screening procedure were sequenced at the Washington University Genome Sequencing Centre (St Louis, USA). The sequence information for this BAC clone was obtained from GenBank: AC192436 (Oa_Bb-484F22).

#### *Computational characterisation of the α- and β-globin clusters in the platypus*

Using sequence information of AC195438, AC203513 and AC192436, genes were predicted by GENSCAN [46] and GenomeScan [47] using default settings. All predicted gene sequences were then subjected to a BLAST search of the translated nucleotide acid (BlastX) and protein (BlastP) databases to confirm their identities.

#### *Promoter analyses*

Transcription factor binding motifs were predicted in the 200 bp promoter region located 5' to the predicted platypus α- and β-like genes and *GBY* by rVista 2.0 [67] using user-defined consensus sequences for 'CACCC', 'CAAT', 'TATA', GATA1 ('WGATAR' [51]) and EKLF ('NGNGTGGGN' [51]). The same criteria were used to predict the same motifs in marsupials (*Didelphis virginiana* [ζ and ψζ': AC139599] and *Sminthopsis macroura* [$\alpha^D$, $\psi\alpha^3$, $\alpha^2$, $\alpha^1$, ω: AC146781; and ε, β: AC148754]) and in humancs [ζ, ψζ', $\alpha^D$, $\psi\alpha^3$, $\alpha^2$, $\alpha^1$: NG_000006; and ε, β: NG_000007] for consistency in comparison.

#### *Confirmation of α1 and α3 by BLAST search and Southern blot*

To confirm that the presence of two almost identical genes ($\alpha^1$ and $\alpha^3$) was real rather than an assembly error, the boundaries (~300 bp) of the homologous regions were investigated by a BLAST search against the platypus WGS database. The raw sequences of best hits were extracted from NCBI GenBank, cleaned and aligned in Sequencher v4.8 (Gene Codes Corporation, Michigan) using default settings.

Southern blotting was also used to verify the presence of $\alpha^1$ and $\alpha^3$ genes. In a 30 μl reaction, 100 μg BAC DNA (Oa_Bb: 131M24, 130N2, 150K14 and 223I12) was digested with 10 U of restriction enzyme,*EcoRV* (Roche, Australia). The reaction was incubated at 37°C for at least 4 hours and separated by electrophoresis on a 0.8% agarose gel overnight at 40 V. The DNA fragments were transferred onto a Hybond N+ (GE Healthcare, Australia) nylon filter overnight by capillary action following the manufacturer's instructions, and cross-linked in 0.4 M NaOH for 20 minutes. These filters were then screened with the platypus $\alpha^1/\alpha^3$ (test) and $\alpha^2$ (control) probes using the standard library screening procedure (above).

#### *RT-PCR analyses*

To remove DNA contamination, RNAs derived from adult male platypus liver, kidney, spleen, testis, brain and lungs were DNase treated using a DNA-*free*™ kit according to the manufacturer's instructions (Applied Biosystems, Australia). Treated RNAs were then reverse transcribed using Superscript III (Invitrogen, Australia) following the manufacturer's instructions. Primers were designed against predicted α- and β- like and *GBY* globin gene sequences using Primer3 [64]. In each case, the region amplified spanned an intron so that the origin of the template (gDNA or cDNA) was immediately obvious. Primer sequences and the expected sizes of amplified cDNA and gDNA bands are shown in Table 2. PCR reactions and cycling conditions were the same as for screening for the platypus β-globin genes, above. The positive bands were directly sequenced by AGRF (Brisbane, Australia) to confirm their identities. The blood contamination in the tested samples had minimal effect on the observed expression pattern, as some tissues (for example, lung and liver) showed no amplification despite containing large quantities of blood.

#### *Phylogenetic analyses*

Phylogenetic analyses were employed to verify the identities of the platypus globin genes and study the evolutionary relationships of the different members of the α- and β-globin gene families. This study was restricted to the coding domains of the α- and β-globin members and the accession numbers of the sequences used are given in the legends of Figures 2 and 3. Phylogenetic analyses were conducted using MP in PAUP* v.4.0b10 [68], and a BI approach using MrBayes v.3.1.2 [69]. Concordance of trees from each of the different methods, bootstrap proportions and posterior probability estimates were used to examine the robustness of nodes.

MP analyses were conducted for the entire coding sequence matrix and after excluding third codon positions using a heuristic search option and default options (TBR branch swapping), with the exception of using random stepwise addition repeated 100 times. Character state optimisation for MP trees used the DELTRAN option. MP

**Table 2: PCR primers used for amplification of the α- and β-like globin genes including *GBY* from the platypus gDNA and cDNA**

| Gene | Forward Primer | Reverse Primer | gDNA (bp) | cDNA (bp) |
|---|---|---|---|---|
| ζ | GGCCGACAAGACCGCAGTCATCTCCC | CCCGATGGCGCTGATGACT | 527 | 190 |
| $\zeta^1$ | TGACCAAAGGCGACAAGACCT | CCCCGATGGCACCGATGACC | 534 | 198 |
| $\alpha^D$ | GAGGCTGTGAAGAACCTGGA | GGTGTACTCCCCTTGCAGAT | 1793 | 153 |
| $\alpha^2$ | TGGCCCACCTCGATGACCTGG | GGGAAGGTGTCTGGCCACC | 289 | 134 |
| $\alpha^1/\alpha^3$ | GCAAGGCCGCCGGTCACGGC | CGCTGTCCATGTCATCGAAGTGCC | 597 | 192 |
| ω | ATTGTGTCCATCTGGGGAAA | GCTTGGCAAAGTTGCTCTTC | 488 | 232 |
| ***GBY*** | CTGGAAACAGGTGTGCAAGA | CTATCTCCGGGGTGTAGCAG | 3202 | 149 |
| ε | ATCTGAGCGCTGAGGAGAAG | GACAGGTTGCCGAAGGAGTCA | 285 | 142 |
| β | CTGTGGGGGAAAGTGAACAT | GGTCAGCACCTTAGCACCAT | 321 | 168 |

bootstrap analyses [70] were carried out using 1000 bootstrap pseudoreplicates, employing a heuristic search option with random stepwise addition.

The program MODELTEST [71] and the Akaike information criterion (AIC) were used to assess the most appropriate model for BI analyses. The MODELTEST analyses were facilitated using the program MrMTgui v1.0 [72]. The MODELTEST analysis was carried out on separate codon positions for α- and β-globin data sets. For α-globin sequences, a general time reversible (GTR) model [73], with a proportion of invariant sites (I) and unequal rates among sites [74], modelled with a gamma distribution (G) was found to be the most appropriate model to use for first and second codon positions, and a GTR+G model was appropriate for third codon positions under the AIC. For β-globin sequences a GTR+I+G model was considered appropriate for first positions, and a GTR+G model was found to be appropriate for second and third codon positions. The MrBayes analysis was carried out applying these different models to each codon position using an unlinked analysis, with default uninformative priors. Four chains were run simultaneously for 2 million generations in two independent runs, sampling trees every 100 generations. The program TRACER (version 1.3; [75]) was used to assess tree and parameter convergence. For both the α-globin and β-globin analyses all effective sample sizes for all parameters were >1297, indicating a sufficient sample of the parameter space had been taken. A burn-in of 2000 trees (equivalent to 200,000 generations) was chosen for each independent run of MrBayes, with a >50% posterior probability consensus tree constructed from the remaining 36,002 trees (18,001 trees each run).

A BI analysis using MrBayes (version 3.1.2) was also carried out using protein sequence data from β-globin genes. A mixed protein model was used, allowing the optimum model of protein evolution to be assessed from a selection of nine fixed-rate models. The optimum model was found to be the Dayhoff model with a posterior probability of 1.0. The analyses were conducted using two million generations in two independent runs, sampling trees every 100 generations. A burn-in of 2,000 trees was used for each run with a 50% consensus tree constructed from the remaining 36,002 trees.

## Authors' contributions

VSP designed and performed most of the experiments and analysed the data. VSP also drafted the main manuscript. SJBC conducted phylogenetic analyses and contributed to the writing of the manuscript. JED helped in designing the experiments and trouble-shooting experiments. BF, TG, WCW and RKW were involved in sequencing the platypus BAC clone (Oa_Bb-484F22). JAMG conceived and supervised the research and contributed to the writing of the manuscript. All authors read and approved the final manuscript.

## Competing interests

The authors declare that they have no competing interests.

## Additional material

### Additional file 1

***Annotation of the platypus α- and β-like and* GBY *globin genes.*** *This table shows the predicted positions of six α-like, ω and* GBY *globin genes in the platypus BAC clone Oa_Bb-131M24 [GenBank: AC203513], two α-like globin genes in BAC clone Oa_Bb-2L7 [GenBank: AC195438], and two β-like globin genes in BAC clone Oa_Bb-484F22 [GenBank: AC192436 reverse direction].*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1741-7007-6-34-S1.doc]

## Acknowledgements

We thank Dr Frank Grützner (The University of Adelaide) and Tim Hore (The Australian National University) for providing platypus tissues and RNA preparations. We also thank Dr Paul Waters for providing computational help with the identification and confirmation of two identical but different genes $\alpha^1$ and $\alpha^3$.

## References


1. Roesner A, Fuchs C, Hankeln T, Burmester T: **A globin gene of ancient evolutionary origin in lower vertebrates: evidence for two distinct globin families in animals.** *Mol Biol Evol* 2005, **22:**12-20.
2. Fuchs C, Burmester T, Hankeln T: **The amphibian globin gene repertoire as revealed by the *Xenopus genome*.** *Cytogenet Genome Res* 2006, **112:**296-306.
3. Dickerson RE, Geis I: *Hemoglobin: Structure, Function, Evolution and Pathology* Menlo Park, CA: Benjamin Cummings; 1983:65-116.
4. Hentschel CC, Kay RM, Williams JG: **Analysis of *Xenopus laevis* globins during development and erythroid cell maturation and the construction of recombinant plasmids containing sequences derived from adult globin mRNA.** *Dev Biol* 1979, **72:**350-363.
5. Jeffreys AJ, Wilson V, Wood D, Simons JP, Kay RM, Williams JG: **Linkage of adult alpha- and beta-globin genes in *X. laevis* and gene duplication by tetraploidization.** *Cell* 1980, **21:**555-564.
6. Kay RM, Harris R, Patient RK, Williams JG: **Molecular cloning of cDNA sequences coding for the major alpha- and beta-globin polypeptides of adult *Xenopus laevis*.** *Nucleic Acids Res* 1980, **8:**2691-2707.
7. Pisano E, Cocca E, Mazzei F, Ghigliotti L, di Prisco G, Detrich HW III, Ozouf-Costaz C: **Mapping of alpha- and beta-globin genes on Antarctic fish chromosomes by fluorescence *in situ* hybridization.** *Chromosome Res* 2003, **11:**633-640.
8. Gillemans N, McMorrow T, Tewari R, Wai AW, Burgtorf C, Drabek D, Ventress N, Langeveld A, Higgs D, Tan-Un K, Grosveld F, Philipsen S: **Functional and comparative analysis of globin loci in pufferfish and humans.** *Blood* 2003, **101:**2842-2849.
9. Dodgson JB, McCune KC, Rusling DJ, Krust A, Engel JD: **Adult chicken alpha-globin genes alpha A and alpha D: no anemic shock alpha-globin exists in domestic chickens.** *Proc Natl Acad Sci USA* 1981, **78:**5998-6002.
10. Engel JD, Dodgson JB: **Analysis of the adult and embryonic chicken globin genes in chromosomal DNA.** *J Biol Chem* 1978, **253:**8239-8246.
11. Engel JD, Dodgson JB: **Analysis of the closely linked adult chicken alpha-globin genes in recombinant DNAs.** *Proc Natl Acad Sci USA* 1980, **77:**2596-2600.
12. Cooper SJ, Wheeler D, De Leo A, Cheng JF, Holland RA, Marshall Graves JA, Hope RM: **The mammalian alphaD-globin gene lineage and a new model for the molecular evolution of alpha-globin gene clusters at the stem of the mammalian radiation.** *Mol Phylogenet Evol* 2006, **38:**439-448.
13. Goh SH, Lee YT, Bhanu NV, Cam MC, Desper R, Martin BM, Moharram R, Gherman RB, Miller JL: **A newly discovered human alpha-globin gene.** *Blood* 2005, **106:**1466-1472.
14. Hoffmann FG, Storz JF: **The alphaD-globin gene originated via duplication of an embryonic alpha-like globin gene in the ancestor of tetrapod vertebrates.** *Mol Biol Evol* 2007, **24:**1982-1990.
15. Hardison RC, Sawada I, Cheng JF, Shen CK, Schmid CW: **A previously undetected pseudogene in the human alpha globin gene cluster.** *Nucleic Acids Res* 1986, **14:**1903-1911.
16. Lauer J, Shen CK, Maniatis T: **The chromosomal arrangement of human alpha-like globin genes: sequence homology and alpha-globin gene deletions.** *Cell* 1980, **20:**119-130.
17. Orkin SH: **The duplicated human alpha globin genes lie close together in cellular DNA.** *Proc Natl Acad Sci USA* 1978, **75:**5950-5954.
18. Proudfoot NJ, Maniatis T: **The structure of a human alpha-globin pseudogene and its relationship to alpha-globin gene duplication.** *Cell* 1980, **21:**537-544.
19. Wakefield MJ, Graves JA: **Marsupials and monotremes sort genome treasures from junk.** *Genome Biol* 2005, **6:**218.
20. Bininda-Emonds OR, Cardillo M, Jones KE, MacPhee RD, Beck RM, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A: **The delayed rise of present-day mammals.** *Nature* 2007, **446:**507-512.
21. McMillan D, Miethke P, Alsop AE, Rens W, O'Brien P, Trifonov V, Veyrunes F, Schatzkamer K, Kremitzki CL, Graves T, Warren W, Grutzner F, Ferguson-Smith MA, Graves JA: **Characterizing the chromosomes of the platypus (*Ornithorhynchus anatinus*).** *Chromosome Res* 2007, **15:**961-974.
22. Grutzner F, Rens W, Tsend-Ayush E, El-Mogharbel N, O'Brien PC, Jones RC, Ferguson-Smith MA, Graves JA: **In the platypus a meiotic chain of ten sex chromosomes shares genes with the bird Z and mammal X chromosomes.** *Nature* 2004, **432:**913-917.
23. Rens W, Grutzner F, O'brien PC, Fairclough H, Graves JA, Ferguson-Smith MA: **Resolution and evolution of the duck-billed platypus karyotype with an X1Y1X2Y2X3Y3X4Y4X5Y5 male sex chromosome constitution.** *Proc Natl Acad Sci USA* 2004, **101:**16257-16261.
24. Warren WC, Hillier LW, Marshall Graves JA, Birney E, Ponting CP, Grutzner F, Belov K, Miller W, Clarke L, Chinwalla AT, Yang SP, Heger A, Locke DP, Miethke P, Waters PD, Veyrunes F, Fulton L, Fulton B, Graves T, Wallis J, Puente XS, Lopez-Otin C, Ordonez GR, Eichler EE, Chen L, Cheng Z, Deakin JE, Alsop A, Thompson K, Kirby P, Papenfuss AT, Wakefield MJ, Olender T, Lancet D, Huttley GA, Smit AF, Pask A, Temple-Smith P, Batzer MA, Walker JA, Konkel MK, Harris RS, Whittington CM, Wong ES, Gemmell NJ, Buschiazzo E, Vargas Jentzsch IM, Merkel A, Schmitz J, Zemann A, Churakov G, Kriegs JO, Brosius J, Murchison EP, Sachidanandam R, Smith C, Hannon GJ, Tsend-Ayush E, McMillan D, Attenborough R, Rens W, Ferguson-Smith M, Lefevre CM, Sharp JA, Nicholas KR, Ray DA, Kube M, Reinhardt R, Pringle TH, Taylor J, Jones RC, Nixon B, Dacheux JL, Niwa H, Sekita Y, Huang X, Stark A, Kheradpour P, Kellis M, Flicek P, Chen Y, Webber C, Hardison R, Nelson J, Hallsworth-Pepin K, Delehaunty K, Markovic C, Minx P, Feng Y, Kremitzki C, Mitreva M, Glasscock J, Wylie T, Wohldmann P, Thiru P, Nhan MN, Pohl CS, Smith SM, Hou S, Renfree MB, Mardis ER, Wilson RK: **Genome analysis of the platypus reveals unique signatures of evolution.** *Nature* 2008, **453:**175-183.
25. Cooper SJ, Hope RM: **Evolution and expression of a beta-like globin gene of the Australian marsupial *Sminthopsis crassicaudata*.** *Proc Natl Acad Sci USA* 1993, **90:**11777-11781.
26. Cooper SJ, Murphy R, Dolman G, Hussey D, Hope RM: **A molecular and evolutionary study of the beta-globin gene family of the Australian marsupial *Sminthopsis crassicaudata*.** *Mol Biol Evol* 1996, **13:**1012-1022.
27. Koop BF, Goodman M: **Evolutionary and developmental aspects of two hemoglobin beta-chain genes (epsilon M and beta M) of opossum.** *Proc Natl Acad Sci USA* 1988, **85(11):**3893-3897.
28. Lee MH, Shroff R, Cooper SJ, Hope R: **Evolution and molecular characterization of a beta-globin gene from the Australian Echidna *Tachyglossus aculeatus* (*Monotremata*).** *Mol Phylogenet Evol* 1999, **12:**205-214.
29. Goodman M, Koop BF, Czelusniak J, Weiss ML: **The eta-globin gene. Its long evolutionary history in the beta-globin gene family of mammals.** *J Mol Biol* 1984, **180:**803-823.
30. Hardison R, Miller W: **Use of long sequence alignments to study the evolution and regulation of mammalian globin gene clusters.** *Mol Biol Evol* 1993, **10:**73-102.
31. Hardison RC: **Comparison of the beta-like globin gene families of rabbits and humans indicates that the gene cluster 5'-epsilon-gamma-delta-beta-3' predates the mammalian radiation.** *Mol Biol Evol* 1984, **1:**390-410.
32. Weatherall DJ: *The Structure, Organization, and Regulation of Human Genes* 3rd edition. Oxford: Oxford University Press; 1991.
33. Goodman M, Czelusniak J, Koop BF, Tagle DA, Slightom JL: **Globins: a case study in molecular phylogeny.** *Cold Spring Harb Symp Quant Biol* 1987, **52:**875-890.
34. Reitman M, Grasso JA, Blumenthal R, Lewit P: **Primary sequence, evolution, and repetitive elements of the *Gallus gallus* (chicken) beta-globin cluster.** *Genomics* 1993, **18:**616-626.
35. Wheeler D, Hope R, Cooper SB, Dolman G, Webb GC, Bottema CD, Gooley AA, Goodman M, Holland RA: **An orphaned mammalian beta-globin gene of ancient evolutionary origin.** *Proc Natl Acad Sci USA* 2001, **98:**1101-1106.
36. Wheeler D, Hope RM, Cooper SJ, Gooley AA, Holland RA: **Linkage of the beta-like omega-globin gene to alpha-like globin genes in an Australian marsupial supports the chromosome duplication model for separation of globin gene clusters.** *J Mol Evol* 2004, **58:**642-652.
37. Holland RA, Gooley AA 2nd: **Characterization of the embryonic globin chains of the marsupial Tammar wallaby, *Macropus eugenii*.** *Eur J Biochem* 1997, **248:**864-871.
38. Andersen NA, Mesch U, Lovell DJ, Nicol SC: **The effects of sex, season and hybernation on haematology and blood viscosity**

**of free-ranging echidnas (*Tachyglossus aculeatus*).** *Can J Zool* 2000, **78:**174-181.
39. Bently PJ, Herreid CF, Schmidt-Nielsen K: **Respiration of a monotreme, the echidna, *Tachyglossus aculeatus*.** *Am J Physiol* 1967, **212:**957-961.
40. Parsons RS, Atwood J, Guiler ER, Heddle RWL: **Comparative studies on the blood of monotremes and marsupials. I. Haematology.** *Comp Biochem Physiol B* 1971, **39(2):**203-208.
41. Whittaker RG, Thompson EO: **Studies on monotreme proteins. V. Amino acid sequence of the alpha-chain of haemoglobin from the platypus, *Ornithorhynchus anantinus*.** *Aust J Biol Sci* 1974, **27:**591-605.
42. Whittaker RG, Thompson EO: **Studies on monotreme proteins. VI. Amino acid sequence of the beta-chain of haemoglobin from the platypus, *Ornithorhynchus anatinus*.** *Aust J Biol Sci* 1975, **28:**353-365.
43. Thompson EO, Fisher WK, Whittaker RG: **Studies on monotreme proteins. 3. Amino acid sequence of the alpha- and beta-globin chains of the minor haemoglobin from the echidna, *Tachyglossus aculeatus aculeatus*.** *Aust J Biol Sci* 1973, **26:**1327-1335.
44. Whittaker RG, Fisher WK, Thompson EO: **Studies on monotreme proteins. I. Amino acid sequence of the beta-chain of haemoglobin from the echidna, *Tachyglossus aculeatus aculeatus*.** *Aust J Biol Sci* 1972, **25:**989-1004.
45. **University of California Santa Cruz (UCSC) Genome Browser** [http://genome.ucsc.edu]
46. Burge C, Karlin S: **Prediction of complete gene structures in human genomic DNA.** *J Mol Biol* 1997, **268:**78-94.
47. Yeh RF, Lim LP, Burge CB: **Computational inference of homologous gene structures in the human genome.** *Genome Res* 2001, **11:**803-816.
48. Motamed K, Bastiani C, Zhang Q, Bailey A, Shen CK: **CACC box and enhancer response of the human embryonic epsilon globin promoter.** *Gene* 1993, **123:**235-240.
49. Kim CG, Sheffery M: **Physical characterization of the purified CCAAT transcription factor, alpha-CP1.** *J Biol Chem* 1990, **265:**13362-13369.
50. Sabath DE, Koehler KM, Yang WQ, Phan V, Wilson J: **DNA-protein interactions in the proximal zeta-globin promoter: identification of novel CCACCC- and CCAAT-binding proteins.** *Blood Cells Mol Dis* 1998, **24:**183-198.
51. Whyatt DJ, deBoer E, Grosveld F: **The two zinc finger-like domains of GATA-1 have different DNA binding specificities.** *EMBO J* 1993, **12:**4993-5005.
52. Gillemans N, Tewari R, Lindeboom F, Rottier R, de Wit T, Wijgerde M, Grosveld F, Philipsen S: **Altered DNA-binding specificity mutants of EKLF and Sp1 show that EKLF is an activator of the beta-globin locus control region *in vivo*.** *Genes Dev* 1998, **12:**2863-2873.
53. Agarwal S, Arya V, Stolle CA, Pradhan M: **A novel Indian beta-thalassemia mutation in the CACCC box of the promoter region.** *Eur J Haematol* 2006, **77:**530-532.
54. Hardison R, Chao KM, Adamkiewicz M, Price D, Jackson J, Zeigler T, Stojanovic N, Miller W: **Positive and negative regulatory elements of the rabbit embryonic epsilon-globin gene revealed by an improved multiple alignment program and functional analysis.** *DNA Seq* 1993, **4:**163-176.
55. Hardison R, Slightom JL, Gumucio DL, Goodman M, Stojanovic N, Miller W: **Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights.** *Gene* 1997, **205:**73-94.
56. Cooper SJ, Wheeler D, Hope RM, Dolman G, Saint KM, Gooley AA, Holland RA: **The alpha-globin gene family of an Australian marsupial, *Macropus eugenii*: the long evolutionary history of the theta-globin gene and its functional status in mammals.** *J Mol Evol* 2005, **60:**653-664.
57. **Ensembl Genome Browser** [http://www.ensembl.org/index.html]
58. Flint J, Tufarelli C, Peden J, Clark K, Daniels RJ, Hardison R, Miller W, Philipsen S, Tan-Un KC, McMorrow T, Frampton J, Alter BP, Frischauf AM, Higgs DR: **Comparative genome analysis delimits a chromosomal domain and identifies key regulatory elements in the alpha globin cluster.** *Hum Mol Genet* 2001, **10:**371-382.
59. Hughes JR, Cheng JF, Ventress N, Prabhakar S, Clark K, Anguita E, De Gobbi M, de Jong P, Rubin E, Higgs DR: **Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences.** *Proc Natl Acad Sci USA* 2005, **102:**9830-9835.
60. Taylor TD, Noguchi H, Totoki Y, Toyoda A, Kuroki Y, Dewar K, Lloyd C, Itoh T, Takeda T, Kim DW, She X, Barlow KF, Bloom T, Bruford E, Chang JL, Cuomo CA, Eichler E, FitzGerald MG, Jaffe DB, LaButti K, Nicol R, Park HS, Seaman C, Sougnez C, Yang X, Zimmer AR, Zody MC, Birren BW, Nusbaum C, Fujiyama A, Hattori M, Rogers J, Lander ES, Sakaki Y: **Human chromosome 11 DNA sequence and analysis including novel gene identification.** *Nature* 2006, **440:**497-500.
61. Bulger M, van Doorninck JH, Saitoh N, Telling A, Farrell C, Bender MA, Felsenfeld G, Axel R, Groudine M: **Conservation of sequence and structure flanking the mouse and human beta-globin loci: the beta-globin genes are embedded within an array of odorant receptor genes.** *Proc Natl Acad Sci USA* 1999, **96:**5129-5134.
62. Opazo JC, Hoffmann FG, Storz JF: **Genomic evidence for independent origins of beta-like globin genes in monotremes and therian mammals.** *Proc Natl Acad Sci USA* 2008, **105:**1590-1595.
63. Hardison RC: **New views of evolution and regulation of vertebrate beta-like globin gene clusters from an orphaned gene in marsupials.** *Proc Natl Acad Sci USA* 2001, **98:**1327-1329.
64. Rozen S, Skaletsky H: **Primer3 on the WWW for general users and for biologist programmers.** *Methods Mol Biol* 2000, **132:**365-386.
65. The ENCODE Project Consortium: **The ENCODE (ENCyclopedia Of DNA Elements) Project.** *Science* 2004, **306:**636-640.
66. **NIH Intramural Sequencing Center (NISC)** [http://www.nisc.nih.gov/]
67. Loots GG, Ovcharenko I: **rVISTA 2.0: evolutionary analysis of transcription factor binding sites.** *Nucleic Acids Res* 2004:W217-W221.
68. Swofford DL: **PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4.0b10.** Sunderland, MA: Sinauer; 2002.
69. Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogeny.** *Bioinformatics* 2001, **17:**754-755.
70. Felsenstein J: **Confidence limits on phylogenies: an approach using the bootstrap.** *Evolution* 1985, **39:**783-791.
71. Posada D, Crandall K: **Modeltest: testing the model of DNA substitution.** *Bioinformatics* 1998, **14:**817-818.
72. **MrMTgui v1.0** [http://www.genedrift.org/mtgui.php]
73. Rodríguez F, Oliver JF, Marín A, Medina JR: **The general stochastic model of nucleotide substitutions.** *J Theor Biol* 1990, **142:**485-501.
74. Yang Z: **Among-site rate variation and its impact on phylogenetic analyses.** *Trends Ecol Evol* 1996, **11:**367-372.
75. Rambaut A, Drummond AJ: **Traver v1.3.** 2004 [http://beast.bio.ed.ac.uk/Tracer].

Publish with **BioMed Central** and every scientist can read your work free of charge

*"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."*

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

BioMedcentral

# Additional File

**Additional File 1: Annotation of the platypus α- and β- like and *GBY* globin genes**

The table shows the predicted positions of six α-like, ω and *GBY* globin genes in the platypus BAC clone Oa_Bb-131M24 [GenBank: AC203513], two α-like globin genes in BAC clone Oa_Bb-2L7 [GenBank: AC195438], and two β-like globin genes in BAC clone Oa_Bb-484F22 [GenBank: AC192436 reverse direction].

| | Exon 1 | Exon 2 | Exon 3 | Poly-A |
|---|---|---|---|---|
| **AC203513** | | | | |
| ζ | 3746 - 3840 | 4178 - 4382 | 4497 - 4625 | 4744 |
| ζ' | 8545 - 8639 | 8976 - 9180 | 9283 - 9411 | 9544 |
| $\alpha^{D}$ | 11992 - 12083 | 13398 - 13602 | 15213 - 15341 | 15418 |
| $\alpha^{3}$ | 20056 - 20147 | 20553 - 20757 | 20909 - 21037 | 21131 |
| $\alpha^{2}$ | 27074 - 27168* | 27889 - 28093 | 28249 - 28377 | 28492 |
| $\alpha^{1}$ | 32437 - 32528 | 32934 - 33138 | 33290 - 33418 | 33512 |
| ω | 39008 - 39099 | 39356 - 39578 | 39690 - 39818 | 39887 |
| ***GBY*** | 48012 - 47915 | 44550 - 44328 | 41277 - 41134 | 40500 |
| **AC195438** | | | | |
| ζ | 137751 - 137845 | 138183 - 138387 | 138502 - 138630 | 138749 |
| ζ' | 142550 - 142644 | 142981 - 143185 | 143288 - 143416 | 143549 |
| **AC192436** | | | | |
| ε | 79553 - 79644 | 79788 - 80010 | 80485 - 80613 | 80709 |
| β | 91660 - 91751 | 91905 - 92127 | 92566 - 92694 | 92765 |

* the $\alpha^{2}$ gene has two sequencing errors in its exon 1 at positions 2 (A->T) and 7 (G->C), which was confirmed by a BLAST search of the platypus WGS database.

# CHAPTER 3: THE EVOLUTIONARY HISTORY OF GLOBIN GENES: INSIGHTS FROM MARSUPIALS AND MONOTREMES

***Authors:***

**Patel, V.S.** and Deakin, J.E.

***Current status of paper (circle as appropriate):***

Planned/ In Preparation/ Submitted/ Under Revision/ Accepted/ Published

***Date paper accepted for publication or anticipated date of publication:***

July 2010

***Name of journal/book:***

Marsupial Genetics and Genomics. Editors: Deakin, J.E., Waters, P.D., and Graves, J.A. M. Springer publishing. p 415-433

***Extent to which research is your own:***

This is a review article in which I provided the major contribution in reviewing a wide literature.

***Your contribution to writing the paper:***

I wrote the draft of the manuscript and prepared figures and incorporated suggestions made by my co-author. I liased with the book editors and responded to their comments

***Comments:***

This manuscript was an invited contribution to a book on 'Marsupial Genetics and Genomics'.

# Chapter 20
# The Evolutionary History of Globin Genes: Insights from Marsupials and Monotremes

**Vidushi S. Patel and Janine E. Deakin**

**Abstract** Haemoglobin, which is required for oxygen transportation in the blood, is encoded by members of the alpha (α) and beta (β) globin gene families. They are highly regulated throughout different stages of development in a tissue-specific manner. The number and type of α- and β-globin genes varies between jawed vertebrates. In teleosts and amphibians α- and β-like globin genes are clustered together, but in birds and mammals these genes form distinct clusters on different chromosomes. This chapter reviews how data from marsupials and monotremes have contributed in discovering novel globin genes, which in turn has clarified how globin genes have evolved throughout amniote evolution. We also provide a detailed view of how the duplication of a single primordial globin gene (about 500 million years ago) along with other subsequent events have resulted in more complex α- and β-globin clusters in extant vertebrates. In addition, how this has impacted the current fate and regulation of α- and β-genes in mammals will also be discussed.

**Keywords** α-globin · β-globin · Evolution · Haemoglobin · Marsupials · Monotremes · Transposition

## 20.1 Introduction

Haemoglobin is vital for the survival of all species, and originated as an iron-complexed protein at about the time when cellular life originated on earth (Hardison, 2001). It is found in all living organisms including prokaryotes, fungi, plants and animals, and its function ranges from catabolic metabolism in bacteria, yeast and algae, to intercellular oxygen transportation in vertebrates (reviewed in Hardison, 1998). All globin genes are thought to have evolved from a single common ancestral globin via a series of gene duplications, which, due to selective pressures and

V.S. Patel (✉)
ARC Centre of Excellence for Kangaroo Genomics, Evolution, Ecology and Genetics, Research School of Biology, The Australian National University, Canberra, ACT 0200, Australia
e-mail: vidushi.patel@anu.edu.au

J.E. Deakin et al. (eds.), *Marsupial Genetics and Genomics*,
DOI 10.1007/978-90-481-9023-2_20, © Springer Science+Business Media B.V. 2010



**Publication (Patel et al. 2010)**

changes in atmospheric oxygen levels, led to the diversity of specialised forms and functions of haemoglobin across all kingdoms of organisms (Hardison, 1998).

In vertebrates, haemoglobin is a major component in blood, but its structure differs from species to species and among isoforms. In agnatha (jawless vertebrates), such as lamprey and hagfish, the haemoglobin molecule is a monomer in the oxygenated form, but changes to heterodimers and heterotetromers in the deoxygenated form (Qiu et al., 2000; Muller et al., 2003). In all gnathostomes (jawed vertebrates), it is always found as a heterotetrameric molecule composed of two alpha (α)-globin and two beta (β)-globin polypeptides, with a heme tightly bound to a pocket in each globin monomer. In highly oxygenated surroundings, i.e. in lungs or gills, the interaction between each globin polypeptide chain allows heme to cooperatively bind to oxygen and unloads it efficiently in the peripheral respiring tissues that need oxygen for efficient metabolism.

Most gnathostomes possess multiple variants of α-like and β-like globin genes that are located in clusters, and upon activation constitute different forms of haemoglobin during different stages of ontogeny. The multiplicity of these genes has the advantage of producing different haemoglobin molecules with distinct oxygen binding affinity during different stages of development and in different tissues. This enables the organism to cope with environmental changes in oxygen tension (Coates, 1975). For example, early expressed haemoglobins in embryos lack cooperative oxygen binding and have a higher oxygen affinity than adult haemoglobin. It allows the embryo to extract more oxygen with greater efficiency from their surroundings or maternal circulation (Hofmann et al., 1997; Brownlie et al., 2003). As the gas exchange structures develop and environmental conditions improve around fishes and amphibians, or placental barriers are reduced in eutherian mammals, these high oxygen affinity haemoglobins are then slowly replaced by lower affinity haemoglobins with greater cooperative binding (Iuchi, 1973; Brittain et al., 1997). Thus, there is a huge demand for coordinating and regulating these α- and β-globin genes during different stages of ontogeny and tissues in gnathostomes for their balanced production in forming proper haemoglobin molecules, which otherwise if altered could lead to haemoglobinopathies such as sickle cell amenia and thalassemia (Wells, 1999).

## 20.2 Unique Globin Properties in Marsupial Newborns

In marsupial embryos and newborns, the red blood cells and haemoglobin properties are quite different to those in older pouch young or adult marsupials. At birth, the red blood cells of marsupials are all nucleated and larger, whereas in more mature pouch young and adults the red blood cells lack nuclei and are smaller (Holland and Gooley, 1997). Haemoglobin in marsupial embryos from species such as the tammar wallaby (*Macropus eugenii*) and brush-tail possum (*Trichosurus vulpecula*) have lower affinity for oxygen than other vertebrate embryos and foetuses (Tibben et al., 1991; Calvert et al., 1994). This allows marsupial embryos to extract less oxygen from maternal circulation (Holland and Gooley, 1997). The lower affinity

for oxygen is really interesting, particularly given that marsupials have a short period of gestation and are born at a very pre-mature stage of development with an underdeveloped respiratory system. In the tammar wallaby neonate, there are four different types of haemoglobin molecules in embryonic blood, which are different from those found in adults (Holland et al., 1998). These molecules are also larger than the adult haemoglobin molecules, containing probably eight sub-units rather than the normal four sub-units (Holland and Gooley, 1997; Holland et al., 1998). These specialised forms of haemoglobin may play a role in the transition from placental respiration in utero to pulmonary respiration in the pouch, since they disappear a few days after birth once the nucleated red blood cells are replaced by adult-like non-nucleated red blood cells (Tyndale-Biscoe, 2005).

These unique properties make studying marsupial globin genes of particular interest. Genomic studies on genes encoding haemoglobin molecules have led to the discovery of novel globin genes in marsupials, which then sparked new interest in determining the evolutionary history of jawed vertebrate α- and β-globin genes.

## 20.3 How Did α- and β-Globin Clusters Evolve in Jawed Vertebrates?

### *20.3.1 History of Globin Gene Evolution*

Globin gene evolution can be traced by studying organisational gene structure and function of α- and β-globin genes in different jawed vertebrate lineages (i.e. from cartilaginous fishes to humans). In cartilaginous fishes such as sharks and ray fishes, although the whole organisational gene structure has not yet been studied, the structure of their haemoglobin molecules is similar to that of other fishes and tetrapods, being composed of two α and two β globin polypeptide chains (Chong et al., 1999; Naoi et al., 2001; Verde et al., 2005). These α and β chains have many structural similarities with human α- and β-globin genes respectively (Chong et al., 1999; Naoi et al., 2001), and therefore the sequence and structure of these chains have been conserved across all jawed vertebrates and evolved from a common ancestral globin gene before fish-tetrapod divergence (~500 MYA).

In teleosts and amphibians, α- and β-globin gene families, each containing a set of differentially expressed genes, are closely linked. The main difference between these two species are that in teleosts the globin genes are linked in a "head to head" orientation (α, 3′–5′; β, 5′–3′), so that they are transcribed in opposite directions (Wagner et al., 1994; Chan et al., 1997; Miyata and Aoki, 1997; Brownlie et al., 2003), whereas in amphibians the globin genes are arranged in a "head to tail" orientation (α, 5′–3′, β, 5′–3′) (Jeffreys et al., 1980; Patient et al., 1980; Hosbach et al., 1983; Fuchs et al., 2006). Also, in the amphibians (*Xenopus laevis* and *X. tropicalis*), a new member of the globin superfamily was recently discovered adjacent (3′) to the α-β cluster. This gene called Globin Y (*GBY*) encoded a *bona fide* functional globin peptide of 156 amino acid and bore sequence hallmarks of a functional respiratory

protein (i.e. contains three-exons/two-introns with conserved splicing sites), but the precise function is currently unknown (Fuchs et al., 2006).

Unlike teleostan fishes and amphibians, birds and eutherian mammals possess two unlinked clusters for α- and β-genes that are located on different chromosomes. The number of globin genes in each of the α- and β-globin clusters are also different. For example, the chicken (*Gallus gallus*) α-globin cluster contains three α-like globin genes (*HBP, HBK, HBA*) located on chromosome 14 (Engel and Dodgson, 1980; Dodgson et al., 1981), whereas their β-globin clusters contain four β-globin genes (*HBR, HBB-T1, HBB-T2, HBE*) located on chromosome 1 (Dolan et al., 1981; Villeponteau and Martinson, 1981) (Fig. 20.1). In eutherian mammals,

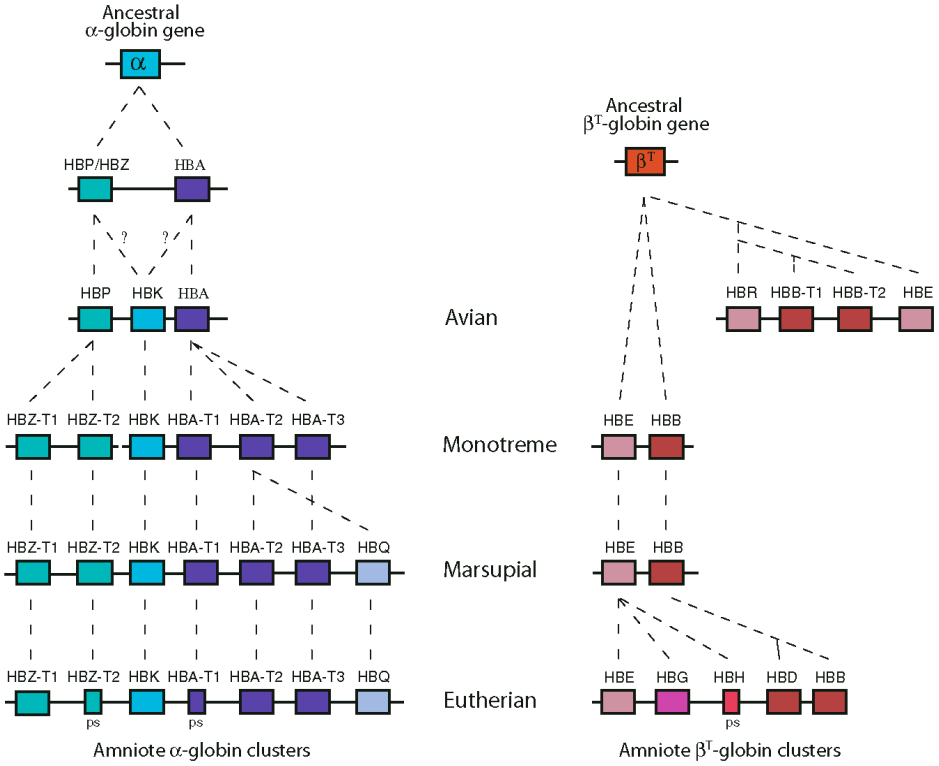



**Fig. 20.1** The gene organisation, structure and evolution of α- and β-like globin genes in different amniote lineages. Globin genes are named according to the new standard nomenclature system specified by Aguileta et al. (2006b), where they should start with *HB* representing haemoglobin and the third letter representing their generic names. For example, *HBA* is for α-globin gene in haemoglobin, *HBB* is for β-globin gene in haemoglobin, E is for epsilon (ε), P for pi (π), K for alphaD ($\alpha^D$), R for rho (ρ), and so forth. The symbol "-T" followed by a number indicates that the gene is a part of a known tandemly duplicated gene block and the number corresponds to the linkage order within the block according to the 5′–3′ orientation. The lowercase letter "ps" at the end of the symbol (in the text) or below the *squares* (in the figure) designates a psuedogene (Aguileta et al., 2006b). This figure not only shows the gene organization in various amniote species, but also how they evolved from a common ancestral α- and $\beta^T$-globin genes and their orthologous relationship

all of these genes are present but their numbers have expanded due to many tandem duplications within the cluster. For example, the human (*Homo sapiens*) α-globin cluster contains seven genes arranged in the order *HBZ-T1, HBZ-T2ps, HBK, HBA-T1ps, HBA-T2, HBA-T3, HBQ* and located on chromosome 16 (Orkin, 1978; Lauer et al., 1980; Proudfoot and Maniatis, 1980; Goh et al., 2005), whereas their β-globin cluster (*HBE, HBG-T1, HBG-T2, HBHps, HBD, HBB*) is located on chromosome 11 (Bernard et al., 1979; Fritsch et al., 1980). Irrespective of their chromosomal locations, the amino acid sequences of α and β-globin genes display at least 50% similarity, with similar structures (three-exons/two-introns) to that of fishes and amphibians, suggesting that the structure of globin genes has been conserved since prior to fish-tetrapod divergence.

There are a lot of sequence and structural similarities between the avian and eutherian α-globin clusters, but not between their β-globin clusters. The genes in the avian α-globin cluster show high sequence similarities to their orthologous counterparts in the mammalian α-globin cluster (Proudfoot et al., 1982). For example, *HBP* from chicken is orthologous to *HBZ* from mammals (Fig. 20.1) with 71% sequence identity at protein level. Moreover, the genes in both clusters are arranged in their order of expression, that is, *HBP/HBZ* are first expressed during embryonic stages, followed by *HBK* during late embryonic/fetal and adult stages, and finally *HBA* is dominantly expressed during adult life (Bruns and Ingram, 1973; Proudfoot et al., 1980). However, these properties are not seen for their β-globin clusters. The alignment score between the avian and mammalian β-globin sequences are poor (Reitman et al., 1993). For instance, the avian *HBE* is as equally similar to its eutherian orthologue (*HBE*) as it is to other members of the eutherian β-globin cluster. The genes are also not arranged in their order of expression in aves (Dodgson et al., 1981) but are in mammals (Lacy et al., 1979; Lauer et al., 1980; Maniatis et al., 1980). These differences are believed to be a result of independent duplications of an ancestral β-globin gene in the avian and mammalian lineages (Proudfoot et al., 1982; Reitman et al., 1993).

The fact that α- and β-genes are closely linked in teleosts and amphibians, their structures are similar to eutherian globin genes, and the eutherian α- and β-globin amino acid sequences are homologous, suggests that these genes evolved by a tandem duplication of the ancestral primordial globin gene some time prior to the divergence of gnathostomes ~500 MYA (Fig. 20.2; Dayhoff et al., 1972; Jeffreys et al., 1980; Goodman et al., 1987). Therefore, the duplicated ancestral proto α-β globin genes were closely linked in fishes and frogs, but became separated in the common ancestor of all amniotes (315 MYA).

Jeffreys et al. (1980) were the first to hypothesize two possible models for the evolution of separate α- and β-globin clusters in amniotes. One model suggested that the separation of these genes occurred by translocation between the α- and β-globin genes, possibly by chromosomal rearrangements after amphibian-amniote divergence, resulting in two unlinked clusters (Pisano et al., 2003). The alternative model suggested that two distinct clusters arose via chromosomal duplication, where the duplicate copies would then have evolved to the present arrangement found in birds and mammals by tandem duplications and silencing of

the linked α- or β-globin genes within each cluster. According to this model, some evolutionary remnants of inactivated globin genes would be present near contemporary amniote globin gene clusters. For instance, a "fossil" β-like globin gene would be present within the amniote α-globin cluster and likewise, a "fossil" α-like globin gene would be present within the amniote β-globin cluster.

### *20.3.2 The Discovery of a Novel Marsupial Globin Gene: Implications for Globin Gene Evolution*

Early work focused on studying globin genes in birds and eutherian mammals leaving a void in our knowledge of globin genes from lineages falling between these two divergent groups of animals. Studies of α- and β-globin genes in marsupials, thus, provided useful information not only about the organization of α- and β-globin clusters, but also led to the discovery of novel globin genes. This has permitted the reconstruction of the ancestral α- and β-globin clusters in the therian ancestor.

Characterisation of marsupial globin genes showed that their α-globin clusters were similar to humans (i.e. *HBZ-T1, HBZ-T2, HBK, HBA-T1, HBA-T2, HBA-T3, HBQ*; Fig. 20.1) and located on chromosome 1 in the tammar wallaby and striped-faced dunnart (*Sminthopsis macroura*) (De Leo et al., 2005; Cooper et al., 2006). The *HBK* (known as $\alpha^{D}$-globin or μ-globin) was previously thought only to be present in birds and reptiles, and lost from the genomes of mammals. However, the discovery of *HBK* in marsupials led to the discovery of *HBK* in eutherian mammals (Goh et al., 2005; Cooper et al., 2006; Hoffmann and Storz, 2007). On the other hand, marsupials have much simpler β-globin clusters than do birds and eutherian mammals, being comprised of only two β-globin genes (*HBE, HBB*; Fig. 20.1) (Koop and Goodman, 1988; Cooper et al., 1996) located on chromosomes 5 and 3 in the tammar wallaby and dunnart respectively (De Leo et al., 2005; Deakin et al., 2008). Since both the α- and β-globin clusters were located on different chromosomes and were orthologous to their respective eutherian α- and β-globin counterparts, it provided support for the Jeffreys et al. (1980) translocation hypothesis where the proto α-β separated from each other by a possible split between them sometime before the radiation of amniotes (>315 MYA).

The recent discovery of a third β-like globin gene called *HBW* in marsupials sparked new interested into determining the evolutionary history of globin genes. The marsupial-specific *HBW* was an extremely interesting discovery since *HBW* was not linked to the main β-globin cluster, but instead was found at the 3′ end of the α-globin cluster (Wheeler et al., 2001, 2004). This "fossil" β-like gene, like other marsupial β-globin genes, was expressed in tammar wallaby neonates and bound with α globin polypeptides to form functional haemoglobin (Holland and Gooley, 1997; Holland et al., 1998). However, the phylogenetic analysis surprisingly indicated that the *HBW* was more closely related to bird β-globin genes rather than mammalian β-globin genes (Wheeler et al., 2001, 2004).

Since *HBW* was seen as an "orphan" (i.e. separated from the other β-globin genes), it provided support for the Jeffreys et al. (1980) chromosome duplication

hypothesis where the whole proto α-β cluster duplicated to another region of the genome and underwent differential gene silencing in different amniote lineages. One of the two globin gene clusters became the β-globin cluster in birds, but the β-globin genes within this cluster became redundant and inactive, and finally deleted in mammals, except for the marsupial *HBW*, which was seen as the last mammalian relic β-globin gene in this cluster (Wheeler et al., 2001, 2004). The other cluster diverged and duplicated to form the β-globin cluster of marsupials and eutherian mammals. Thus, under this model, the avian β-globin genes were orthologous to marsupial *HBW*, and both were paralogous to the mammalian β globin genes (Wheeler et al., 2001, 2004), contradicting previous claims that the β-globin clusters of birds and mammals were orthologous.

The chromosome duplication hypothesis answered some questions but ignited new ones. It made it clear why a "fossil" and "orphaned" β-like *HBW* gene was found adjacent (3′) to the α-globin cluster. This also partially explained why chicken β-globin genes did not show high sequence similarities to mammalian β-globin genes (Reitman et al., 1993). However, the *HBW* gene was not a pseudogene, but a transcriptionally active gene that encoded a functional haemoglobin molecule (reviewed in Hardison, 2008). Also, no evidence has been found for a fossil or functional α-like globin gene adjacent to any amniote β-globin cluster, nor have orthologues of the chicken β-globin gene cluster and marsupial *HBW* gene been found in eutherian mammals. It also failed to explain why the regions spanning the duplicated globin clusters are so different, i.e. the mammalian α-globin clusters reside in a G+C-rich region that contains a CpG island, whereas their β-globin clusters are in an A+T-rich region (reviewed in Hardison, 1998). If the whole α-β cluster duplicated, then one would expect to see some flanking genes and regulatory regions common to both clusters. Therefore, the model proposed by Wheeler et al. (2001, 2004) explained the presence of the *HBW* gene beside the marsupial α-globin cluster, but failed to explain the differences between α- and β-globin clusters in amniotes.

### *20.3.3 A New Model for Globin Gene Evolution: Insights from Monotremes*

The discovery of the marsupial *HBW* gene and the poor explanation of globin evolution by chromosome duplication sparked further interest in solving the mystery of globin gene evolution by looking at the most basal lineage of mammals, the monotremes. This particular mammalian lineage has unique characteristics distinct from other mammals, such being the only egg-laying mammals.

Studies of globin genes in the platypus (*Ornithorhynchus anatinus*) revealed that, like marsupials, the α-globin cluster contains the *HBW* gene, with the entire cluster consisting of a similar arrangement to other therian mammals i.e. *HBZ-T1, HBZ-T2, HBK, HBA-T1, HBA-T2, HBA-T3, HBW, GBY* (Fig. 20.1; Patel et al., 2008). The only two differences are that it does not possess *HBQ* between *HBA-T3* and *HBW*, suggesting that it evolved by a duplication of *HBA-T2* after the divergence of

monotremes and therian mammals (Fig. 20.1). The platypus α-globin cluster also contained a homolog of *GBY*, a gene recently discovered in amphibians by Fuchs et al. (2006). The presence of *GBY* in both amphibians and monotremes at a similar position within the cluster indicated that it was present in the last common ancestor of tetrapods (>354 MYA).

Furthermore, like marsupials, the platypus had three β-globin genes; two in the main β-globin cluster (*HBE*, *HBB*; Fig. 20.1) located on chromosome 2, and the third (*HBW*) residing adjacent (3′) to the α-globin cluster on chromosome 21 (Opazo et al., 2008; Patel et al., 2008). The finding of *HBW* in monotremes was significant but the phylogenetic analyses presented by Patel et al. (2008) was different to that of Wheeler et al. (2001, 2004). The *HBW* lineage grouped separately from the bird and other mammalian β-globin genes (Patel et al., 2008) (Fig. 20.2). The position of *HBW* in the phlogenetic tree suggested that the *HBW* lineage was more ancient than other β-globin genes in the avian and mammalian β-globin clusters, and that the avian and mammalian β-globin genes were orthologous to each other (Aguileta et al., 2006a; Patel et al., 2008), but paralogous to fish and amphibian β-globin genes and to the marsupial and monotreme β-like (*HBW*) gene.

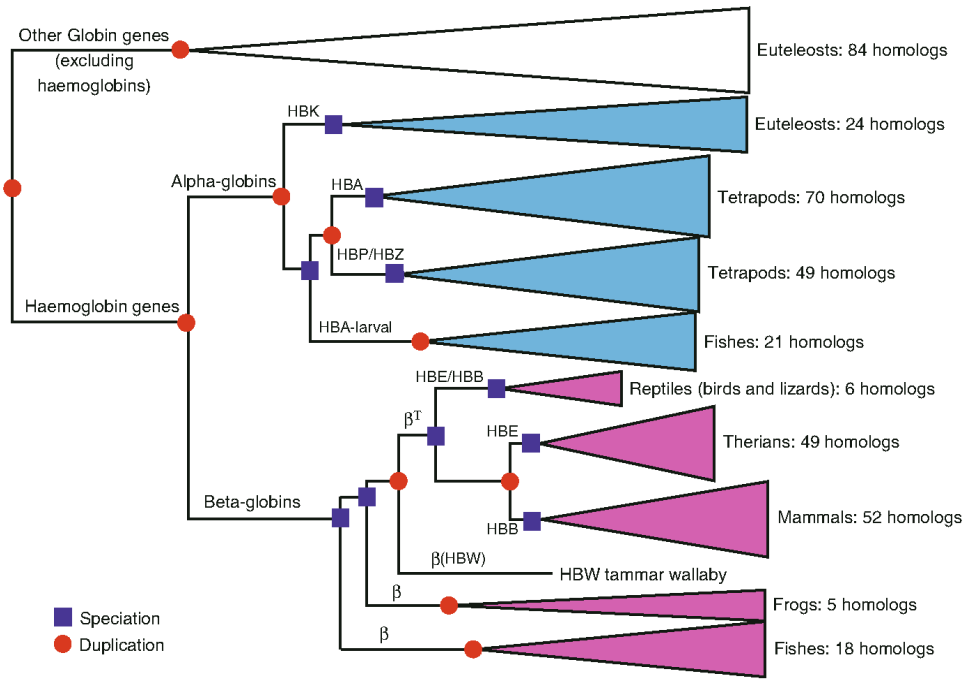



**Fig. 20.2** The phylogeny of all globin genes. The globin phylogeny shows how an ancestral globin gene duplicated into ancestral α- and β-globin genes, followed by further duplications and divergences in various jawed vertebrates. This "Gene Tree" was extracted from Ensembl v56 (15 September 2009; Vilella et al., 2009)

Further evidence for this came through flanking analysis of α- and β-globin clusters in other vertebrates. The avian and mammalian α-globin clusters were flanked by *MPG* and *C16orf35* on the 5′ end, and *GBY* and *LUC7L* on the 3′ end. These 5′ flanking markers were also present adjacent to the α-β cluster of teleosts and amphibians (including GBY), supporting its orthology to the α-globin cluster of all amniotes. Therefore, this ancient α-β cluster (and flanking regions) has been conserved for more than 500 MYA, since before the divergence of gnathostomes (Patel et al., 2008).

After the evolution of gnathostomes and the duplication of an ancestral single primordial globin gene into the ancestral proto α-β globin genes (Fig. 20.3), the proto-α globin gene underwent further tandem duplications, subsequent divergence and specialisation of embryonic and adult globin complexes in different vertebrate lineages. It first duplicated sometime before the teleost-tetrapod divergence more than 450 MYA into the progenitors of embryonic globin genes *HBP/HBZ* and adult globin genes *HBA* (Czelusniak et al., 1982; Goodman et al., 1987; Fig. 20.1). In teleosts and amphibians, further tandem duplications, followed by sequence divergence, expanded the embryonic and adult α- and β-globin genes. Then after amphibian-amniote divergence, there was an additional tandem duplication of either *HBP/HBZ* or *HBA* to result in the *HBK* lineage (Fig. 20.1), since all three genes are present in aves (Engel and Dodgson, 1978; Dodgson et al., 1981; Alev et al., 2009) and mammals (Goh et al., 2005; Cooper et al., 2006; Hoffmann and Storz, 2007). The order, timing and origin of these duplications are still unknown. For example, *HBK* may have evolved by duplication of an adult *HBA* (Cooper et al., 2006) or of an embryonic *HBP/HBZ* and this duplication may have occurred before (and not after) tetrapod divergence with *HBK* lost secondarily in amphibians (Hoffmann and Storz, 2007).

After the divergence of birds and mammals, there were further tandem duplications of embryonic and adult α-like globin genes to produce a six-gene cluster (*HBZ-T1, HBZ-T2, HBK, HBA-T1, HBA-T2, HBA-T3*; Fig. 20.1) in monotremes, with ongoing gene conversion events homogenising the two embryonic (*HBZ-T1* and *HBZ-T2*) and three adult (*HBA-T1,2,3*) genes (Patel et al., 2008). There was an additional tandem duplication of *HBA-T2* to generate *HBQ* after monotreme and therian divergence, thus giving rise to a seven-gene cluster (*HBZ-T1, HBZ-T2, HBK, HBA-T1, HBA-T2, HBA-T3, HBQ*; Fig. 20.1) in marsupial (Cooper et al., 2006) and eutherian mammals (Orkin, 1978; Lauer et al., 1980; Proudfoot and Maniatis, 1980; Goh et al., 2005).

The amniote β-globin cluster resides in a completely different genomic context to that of the α-globin cluster, sharing none of its flanking markers with the amniote α-globin cluster, or even the α-β cluster of teleosts and amphibians. Instead, they are surrounded by many olfactory receptor (*OR*) genes (Bulger et al., 1999). Since *OR* genes are a huge superfamily in vertebrates (reviewed in Delbridge et al., Chapter 21; Niimura and Nei, 2005), it was necessary to look at more distant flanking regions for single-copy genes to make genomic context comparisons. In this case, Patel et al. (2008) used the markers *ILK* and *CCKBR*, which were located 5′ of the β-globin cluster, and *RRM1* that was located 3′. These three flanking markers were

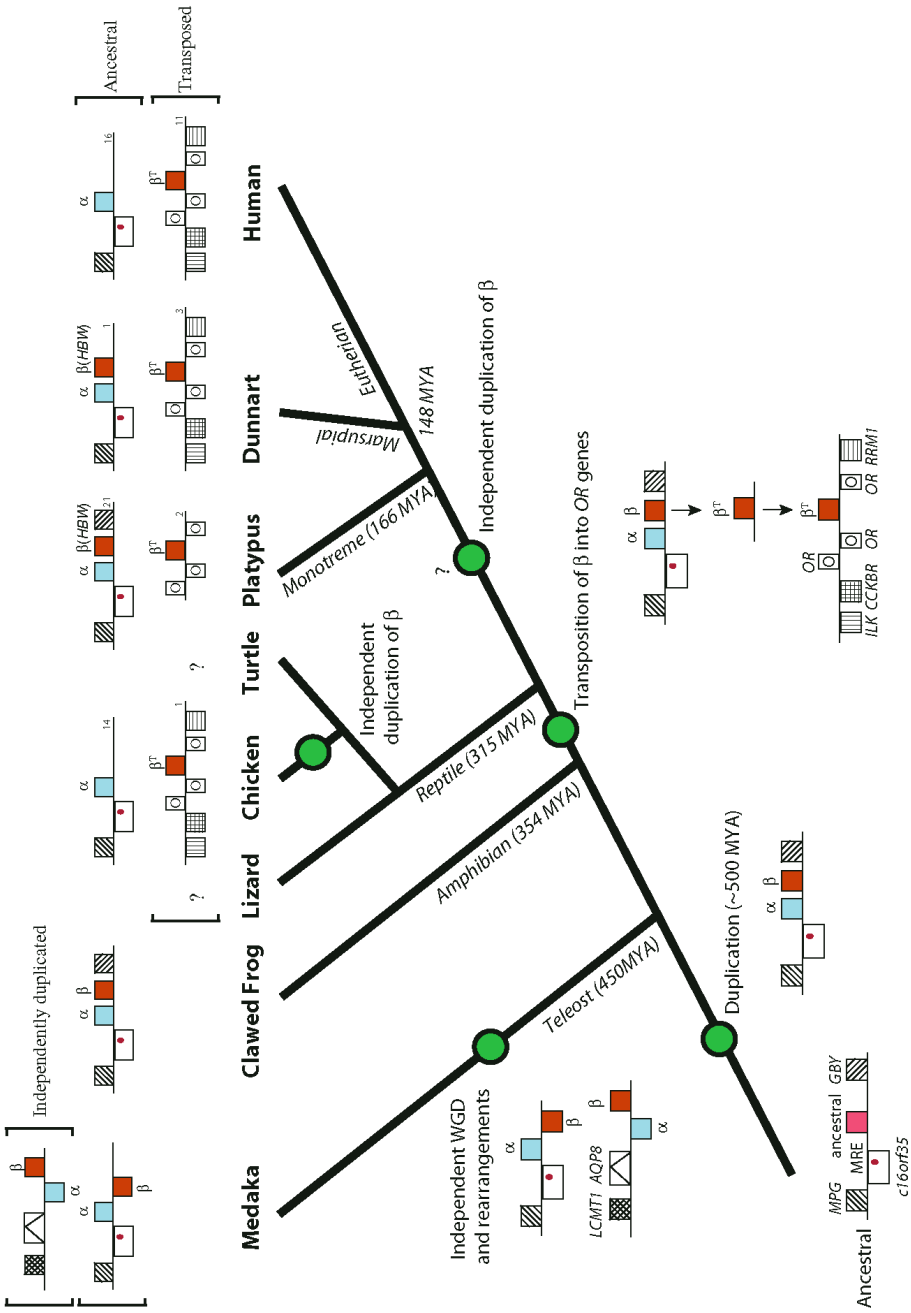



**Fig. 20.3** (continued)

always present in their correct orientation near the avian and mammalian β-globin clusters, but not near to any teleost and amphibian α-β globin clusters, indicating that the amniote β-globin cluster evolved recently after the amphibians and amniotes diverged (Proudfoot et al., 1982; Patel et al., 2008; Hardison, 2008). A copy of an ancestral β-like globin gene was transposed (henceforth referred to as $\beta^T$-globin gene) from the ancient α-β globin cluster and inserted into the array of *OR* genes flanked by *ILK*, *CCKBR* and *ILK* before avian-mammalian divergence (315 MYA), which then duplicated and diverged further to result in modern β-globin clusters (henceforth referred to as $\beta^T$-globin clusters) in contemporary reptilian, avian and mammalian species (Fig. 20.1; Patel et al., 2008).

In the avian lineage, the $\beta^T$-globin gene tandemly duplicated independent to that of mammals giving rise to the four-gene cluster *HBP, HBB-T1, HBB-T2, HBE* (Fig. 20.1), followed by some gene conversions and rearrangements (Czelusniak et al., 1982; Reitman et al., 1993). The $\beta^T$-globin gene also tamdemly duplicated independently in the mammalian lineage, but the number and timing of duplications are still debatable at this stage since there is an uncertainty in the identification of the monotremes *HBE*. Although the platypus possesses two genes in their $\beta^T$-globin cluster (*HBE, HBB*) just like marsupials, its *HBE* gene phylogenetically grouped closely to adult *HBB* genes rather than other mammalian embryonic *HBE* genes (Opazo et al., 2008; Patel et al., 2008). Since platypus *HBE* was located 5′ to *HBB* and had an expression profile similar to other embryonic α-like globin genes (*HBZ-T1* and *HBZ-T2*), Patel et al. (2008) argued that the $\beta^T$-globin gene duplicated into progenitors of embryonic *HBE* and adult *HBB* just prior to monotreme-therian divergence (Fig. 20.1), but the ongoing gene conversions homogenised the platypus *HBE* to group with monotreme adult *HBB*.

Alternatively, Opazo et al. (2008) suggested that there had been two independent duplications of the $\beta^T$-globin gene in the mammalian lineage; one in the monotreme lineage and the other after monotreme-therian divergence, thereby resulting in a total of three independent duplications of the ancestral $\beta^T$-globin gene in amniotes. However, it is not clearly understood how three independent duplications in different lineages would have resulted in genes having a similar pattern of expression during ontogeny. This perhaps could be a consequence of convergent evolution. In either

**Fig. 20.3** Recent model for the evolution of α- and β-globin clusters in vertebrate lineages. The *circles* drawn on the phylogeny tree indicate major events that have occurred during the evolution of globin genes (i.e. from a duplication of a single primordial globin gene to the transposition event followed by complex tandem/independent duplications and subsequent divergence in different lineages). On the *top panel* of each species are the appearance of the clusters and the flanking genes in them. Note: Generic names (α and β) are used in this context to represent the α and β globin clusters and not the individual genes contained in them. The only exception is *HBW*, which specifically represents the β-like globin gene *HBW*, adjacent to the monotreme and marsupial α-globin clusters. The $\beta^T$ represents the transposed β-globin gene found within a cluster of olfactory genes (marked as "O" inside the *box*). For individual genes contained in the α, β and $\beta^T$ globin clusters of various gnathosomes, refer to the text and Fig. 20.1

case, further research into the gene structure and function of β-like globin genes in other birds, reptiles and monotremes would clarify this evolutionary step.

Data from marsupials gave strong evidence that the duplication of the $\beta^T$-globin gene into progenitors of embryonic *HBE* and adult *HBB* occurred at least before the divergence of marsupials and eutherian mammals since these two genes are present in all therian mammals studied to date (Koop and Goodman, 1988; Cooper and Hope, 1993; Cooper et al., 1996; De Leo et al., 2005). After marsupial-eutherian divergence, the progenitors of *HBE* and *HBB* tandemly duplicated to produce a five-gene ancestral cluster (*HBE, HBG, HBH, HBD, HBB*; Fig. 20.1) in eutherian mammals (Goodman et al., 1984). In some mammals, additional events such as block duplications, gene conversions and transpositions have resulted in more complex β-globin clusters.

### *20.3.4 What Was the Mechanism Behind Transposition?*

The evolution of the $\beta^T$-globin cluster in amniotes was most likely due to a single transposition event that moved a copy of the β-globin gene to a new site containing *OR* genes. The exact mechanism for this recruitment remains unclear at this stage but it is likely to have involved transposable elements for the gene's survival. It also seems possible that the region in which the $\beta^T$-globin cluster became embedded in amniotes was a "hot-spot" for transposition events. In marsupials, such as the tammar wallaby and South American opossum (*Monodelphis domestica*), the *FTSJ1* gene, found on the X chromosome in humans, has also transposed into this region beside the $\beta^T$-globin cluster on opossum chromosome 4 and tammar wallaby chromosome 5 (Deakin et al., 2008). Therefore, this region appears to have recruited genes from various parts of the genome.

## 20.4 Regulation of α- and β-Globin Genes

The switching of the α and β globin genes at various developmental stages is common across gnathostomes, indicating a highly conserved regulatory system. The α-globin clusters of birds and therian mammals and the $\beta^T$-globin clusters of therian mammals have distal cis-regulatory elements called the major regulatory region (MRE) located in the intron of *C16orf35* for α-globin clusters (Higgs et al., 1990), and locus control region (LCR) for $\beta^T$-globin clusters (Forrester et al., 1986; Grosveld et al, 1987). These regulatory regions contain conserved binding sites for transcription factors that control the high level, copy number-dependent, position-independent expression of associated globin gene in transgenic mice (Grosveld et al., 1987; Higgs et al., 1990).

Both these regulatory regions differ from each other in many ways: the MRE is about 0.4 kb long, contains only one DNase hypersensitive site and is located about 20–60 kb upstream of the α-globin loci; whereas the LCR comprises five DNase hypersensitive sites (HS 1–5), is located about 6–20 kb upstream of the

$\beta^T$-globin locus and is about 20–25 kb long. Each HS site in the LCR has a core sequence of ~250 bp and contain binding sites for the transcription factors (Stamatoyannopoulos, 2005). The $\beta^T$-globin LCR is also required for tissue-specific chromosomal domain opening, whereas no such function has been defined for α-globin MRE (reviewed in Hardison et al., 1997; Higgs and Wood, 2008). The only similarity between them is that α-globin MRE shares many of the structural features of the HS2 site of the $\beta^T$-globin LCR (Jarman et al., 1991). Furthermore, the tissue-specific and developmental-specific expression of α- and β-like globin genes are not only controlled by MRE and LCR, but also by other promoters and local enhancers located near them (Stamatoyannopoulos, 1991).

The α-globin MRE is well conserved in amniotes but the $\beta^T$-globin LCR is not. The MRE sequences of eutherian mammals show some similarities to those of marsupials (Hughes et al., 2005) and birds, and even to fish, acting as active enhancers (Flint et al., 2001). Although MRE has not been studied in monotremes to date, it is expected that their MRE would be conserved with other mammalian and avian MREs. It has therefore been conserved since the evolution of α-globin genes, prior to the divergence of fish and tetrapods (>450 MYA).

On the other hand, the LCR of eutherian mammals shows conservation only with marsupials (De Leo et al., 2005) and not chickens (Reitman et al., 1993). Data from marsupials show that, although they contain a much simpler $\beta^T$-globin cluster than eutherian mammals, their LCR (HS 1–5) is equivalent to that of eutherians. It therefore supports the holocomplex theory that the LCR acts as an integral unit to regulate the β-globin genes, by activating the transcriptional apparatus at the globin gene promoters (reviewed in Engel and Tanimoto, 2000; De Leo et al., 2005). The conservation of these regulatory regions across taxa indicates that their function has also been conserved, even in marsupials whose haemoglobin molecules have different properties in embryos and newborns. Chickens, on the other hand, although they have four HS sites located upstream of their β-globin loci, they do not have major enhancing activities to confer copy number-dependent and high-expression of the β-globin gene in transgenic mice nor the sequences show any conservation with therian HS sites (Reitman and Felsenfeld, 1990; Reitman et al., 1993; Abruzzo and Reitman, 1994; Mason et al., 1995). Instead, chickens have a strong tissue-specific enhancer located within its β-globin loci called $\beta^A/\varepsilon$ enhancer (Choi and Engel, 1986; Hesse et al., 1986) that together with upstream HS sites and local promoters regulate the high expression of β-globin genes in them (Mason et al., 1995). The lack of conservation between the chicken and therian β-globin regulatory regions (Reitman et al., 1993; Abruzzo and Reitman, 1994) suggests that they evolved independently in the avian and mammalian lineages.

### *20.4.1 How Did Transposition Affect the Regulation of the Globin Clusters?*

In amniotes, the α- and β-globin genes are expressed specifically in erythroid cells and require a complex coordinated regulatory network for the balanced expression

of α and β globin genes to form functional haemoglobin. Given that a $\beta^T$-globin gene transposed into a region containing *OR* genes that are expressed only in nasal epithelium, raises the question of how has it maintained its erythroid-specific expression. All amniote species studied to date (except monotremes) have a *cis*-regulatory element (LCR for mammals and $\beta^A/\varepsilon$ enhancer for chickens) regulating the expression of genes in their $\beta^T$-globin clusters. How this has evolved is a challenging question. Perhaps during transposition of the $\beta^T$-globin gene sufficient regulatory sequences from the α-globin MRE were also transposed, which subsequently duplicated and diverged further in amniotes, thus making it quite different to the α-globin MRE. Alternatively, amniotes evolved their own regulatory regions for $\beta^T$-globin loci after the transposition event followed by duplications and divergence (reviewed in Hardison, 2008).

In either case, it would be predicted that $\beta^T$-globin regulatory regions of aves and mammals would be conserved but indeed they are not. Eutherian LCR share similarities with marsupials only (De Leo et al., 2005), but not with aves (Reitman et al., 1993). The status of the monotreme LCR is debateable at this stage. De Leo et al. (2005) claim to have isolated the platypus HS2 and HS3 sites by amplification using the primers that have been conserved with all other therian mammals. According to De Leo et al. (2005), the sequences of these two hypersensitive sites show high similarities to other marsupial and eutherian HS sites but not to chicken HS sites. Now that the platypus genome assembly (OrnAna1 – March 2007; http://genome.ucsc.edu/cgi-bin/hgGateway) and the bacterial artificial chromosome clones containing the $\beta^T$-globin cluster (Patel et al., 2008) have been sequenced, one would assume that these platypus HS2 and HS3 sites would clearly be present. However, the HS2 and HS3 sequences isolated by De Leo et al. (2005) could not be found in the proximity of the platypus $\beta^T$-globin cluster, or anywhere else in the platypus genome assembly (OrnAna1). Even attempts to re-isolate these two regions using the same primer sequences and the same method described by De Leo et al. (2005) have failed to amplify any fragments (Patel et al., in preparation). These recent results suggest that there was potentially a contamination issue with the experiments described in De Leo et al. (2005). It is therefore, uncertain if platypus does contain conserved HS2 and HS3 sites or contains a different LCR region that controls the expression of $\beta^T$-globin genes.

Given that the transposition event occurred before avian-mammalian divergence (315 MYA), it is possible that the regulatory regions would have diverged significantly, resulting in the low similarity observed between aves and mammals (reviewed in Hardison, 2008). Alternatively, since there has been independent duplications of the $\beta^T$-globin gene in avian and mammalian lineages, it is equally likely that their regulatory regions would have evolved independently with them as well, thereby resulting in different HS sites, enhancers and promoters. Further research into the LCR of monotremes and reptilian species would be useful in clarifying the evolution of $\beta^T$-globin regulatory regions in amniotes.

## 20.5 Unsolved Questions and Future Work

In conclusion, many studies into the gene structures of α- and β-globin clusters in various vertebrates have provided clues about their evolution and regulation. It wasn't until the discovery of the marsupial *HBW* gene (Wheeler et al., 2001, 2004), and the subsequent genomic context analysis in monotremes and other species (Patel et al., 2008), that a simplified model for globin gene evolution was proposed. In the ancestor of jawed vertebrates (500 MYA), a single primordial globin gene duplicated into proto α-β globin genes linked together, followed by further tandem duplications, subsequent sequence divergence and specialisation to result in the modern α-β globin cluster in contemporary teleosts and amphibians, and the α-globin cluster in all amniotes (Fig. 20.3).

In amniotes, a distinct and unlinked β-globin cluster is found on a separate chromosome from that of α-globin cluster. Of all proposed models, the transposition model proposed by Patel et al. (2008) provided the most parsimonious explanation of how a copy of β-globin gene, was transposed into a sea of olfactory receptor genes before the evolution of amniotes. The transposed $\beta^T$-globin gene duplicated and diverged independently in contemporary aves and mammals to form a dominant $\beta^T$-globin cluster. As a consequence, the avian and mammalian $\beta^T$-globin clusters are quite different to each other. Furthermore, this transposition event also explained why both α- and $\beta^T$-globin clusters in amniotes were so different to each other in respect to their flanking regions, regulatory regions and locations. Research into marsupials and monotremes globin genes have been instrumental in this study, as it has clarified the timing of when many globin genes arose in mammals.

We have come a step closer to understanding globin genes, their evolution and regulation in jawed vertebrates. Based on the studied species, models are proposed to gain insight into what might have happened over the course of evolution. However, all models, including the transposition model demonstrated here, need further validation from other species that have not yet been studied. Sequencing the genomes of various species, including the Australian model tammar wallaby, have advanced and could facilitate research into globin studies. The genome of the tammar wallaby could answer questions such as: (1) Does it contain *GBY*? (2) What are the structures of the MRE and LCR of its α- and $\beta^T$-globin clusters respectively? (3) and does the region in close proximity to the $\beta^T$-globin cluster show signs of transposable elements? The data from marsupials also need to be compared to other species such as monotremes and reptiles to provide constructive evaluation of the data, and potentially answer how and why the transposition event took place and the consequences of such movement.

In addition, more intensive research into the expression and regulation of globin genes in marsupials, especially in the embryos and newborns is needed. This would perhaps hold the answer to why some marsupial embryos and newborns have unique haemoglobin properties compared to other amniote embryos. Research into globin genes in marsupials has been, and will continue to be, very valuable not only to study globin gene evolution, but also to study their unique properties and regulation.

**Acknowledgments** We thank Hardip Patel for his assistance with drawing Fig. 20.2.

## References

Abruzzo LV, Reitman M (1994) Enhancer activity of upstream hypersensitive site 2 of the chicken beta-globin cluster is mediated by GATA sites. J Biol Chem 269:32565–32571.

Aguileta G, Bielawski JP, Yang Z (2006a) Evolutionary rate variation among vertebrate beta globin genes: implications for dating gene family duplication events. Gene 380:21–29.

Aguileta G, Bielawski JP, Yang Z (2006b) Proposed standard nomenclature for the alpha- and beta-globin gene families. Genes Genet Syst 81:367–371.

Alev C, Shinmyozu K, McIntyre BA, et al. (2009) Genomic organization of zebra finch alpha and beta globin genes and their expression in primitive and definitive blood in comparison with globins in chicken. Dev Genes Evol 219:353–360.

Bernard R, Little PFR, Annison G, et al. (1979) Structure of the human gammaG-gammaA-delta-beta-globin gene locus. Proc Natl Acad Sci U S A 76:4827–4831.

Brittain T, Hofmann OM, Watmough NJ, et al. (1997) A two-state analysis of co-operative oxygen binding in the three human embryonic haemoglobins. Biochem J 326:299–303.

Brownlie A, Hersey C, Oates AC, et al. (2003) Characterization of embryonic globin genes of the zebrafish. Dev Biol 255:48–61.

Bruns GA, Ingram VM (1973) Erythropoiesis in the developing chick embryo. Dev Biol 30:455–459.

Bulger M, van Doorninck JH, Saitoh N, et al. (1999) Conservation of sequence and structure flanking the mouse and human beta-globin loci: the beta-globin genes are embedded within an array of odorant receptor genes. Proc Natl Acad Sci U S A 96:5129–5134.

Calvert SJ, Holland RAB, Gemmell RT (1994) Respiratory Properties of the Neonatal Blood of the Common Brushtail Possum (*Trichosurus vulpecula*). Physiol Zool 67:407–417.

Chan FY, Robinson J, Brownlie A, et al. (1997) Characterization of adult alpha- and beta-globin genes in the zebrafish. Blood 89:688–700.

Choi OR, Engel JD (1986) A 3′ enhancer is required for temporal and tissue-specific transcriptional activation of the chicken adult beta-globin gene. Nature 323:731–734.

Chong KT, Miyazaki G, Morimoto H, et al. (1999) Structures of the deoxy and CO forms of haemoglobin from *Dasyatis akajei*, a cartilaginous fish. Acta Crystallogr D Biol Crystallogr 55:1291–1300.

Coates ML (1975) Hemoglobin function in the vertebrates: an evolutionary model. J Mol Evol 6:285–307.

Cooper SJ, Hope RM (1993) Evolution and expression of a beta-like globin gene of the Australian marsupial *Sminthopsis crassicaudata*. Proc Natl Acad Sci U S A 90:11777–11781.

Cooper SJ, Murphy R, Dolman G, et al. (1996) A molecular and evolutionary study of the beta-globin gene family of the Australian marsupial *Sminthopsis crassicaudata*. Mol Biol Evol 13:1012–1022.

Cooper SJ, Wheeler D, De Leo A, et al. (2006) The mammalian alphaD-globin gene lineage and a new model for the molecular evolution of alpha-globin gene clusters at the stem of the mammalian radiation. Mol Phylogenet Evol 38:439–448.

Czelusniak J, Goodman M, Hewett-Emmett D, et al. (1982) Phylogenetic origins and adaptive evolution of avian and mammalian haemoglobin genes. Nature 298:297–300.

Dayhoff MO, Hunt LT, McLaughlin PJ, et al. (1972) Gene duplications in evolution. In: Dayhoff MO (ed) Atlas of Protein Sequence and Structure. National Biomedical Research Foundation, Washington, DC, pp. 17–30.

De Leo AA, Wheeler D, Lefevre C, et al. (2005) Sequencing and mapping hemoglobin gene clusters in the Australian model dasyurid marsupial *Sminthopsis macroura*. Cytogenet Genome Res 108:333–341.

Deakin JE, Koina E, Waters PD, et al. (2008) Physical map of two tammar wallaby chromosomes: a strategy for mapping in non-model mammals. Chromosome Res 16:1159–1175.

Dodgson JB, McCune KC, Rusling DJ, et al. (1981) Adult chicken alpha-globin genes alpha A and alpha D: no anemic shock alpha-globin exists in domestic chickens. Proc Natl Acad Sci U S A 78:5998–6002.

Dolan M, Sugarman BJ, Dodgson JB, et al. (1981) Chromosomal arrangement of the chicken beta-type globin genes. Cell 24:669–677.

Engel JD, Dodgson JB (1978) Analysis of the adult and embryonic chicken globin genes in chromosomal DNA. J Biol Chem 253:8239–8246.

Engel JD, Dodgson JB (1980) Analysis of the closely linked adult chicken alpha-globin genes in recombinant DNAs. Proc Natl Acad Sci U S A 77:2596–2600.

Engel JD, Tanimoto K (2000) Looping, linking, and chromatin activity: new insights into beta-globin locus regulation. Cell 100:499–502.

Flint J, Tufarelli C, Peden J, et al. (2001) Comparative genome analysis delimits a chromosomal domain and identifies key regulatory elements in the alpha globin cluster. Hum Mol Genet 10:371–382.

Forrester WC, Thompson C, Elder JT, et al. (1986) A developmentally stable chromatin structure in the human beta-globin gene cluster. Proc Natl Acad Sci U S A 83:1359–1363.

Fritsch EF, Lawn RM, Maniatis T (1980) Molecular cloning and characterization of the human beta-like globin gene cluster. Cell 19:959–972.

Fuchs C, Burmester T, Hankeln T (2006) The amphibian globin gene repertoire as revealed by the *Xenopus* genome. Cytogenet Genome Res 112:296–306.

Goh SH, Lee YT, Bhanu NV, et al. (2005) A newly discovered human alpha-globin gene. Blood 106:1466–1472.

Goodman M, Koop BF, Czelusniak J, et al. (1984) The eta-globin gene. Its long evolutionary history in the beta-globin gene family of mammals. J Mol Biol 180:803–823.

Goodman M, Czelusniak J, Koop BF, et al. (1987) Globins: a case study in molecular phylogeny. Cold Spring Harb Symp Quant Biol 52:875–890.

Grosveld F, van Assendelft GB, Greaves DR, et al. (1987) Position-independent, high-level expression of the human beta-globin gene in transgenic mice. Cell 51:975–985.

Hardison R (1998) Hemoglobins from bacteria to man: evolution of different patterns of gene expression. J Exp Biol 201:1099–1117.

Hardison R, Slightom JL, Gumucio DL, et al. (1997) Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. Gene 205:73–94.

Hardison RC (2001) Organization, evolution and regulation of the globin genes. In: Steinberg MC, Forget BG, Higgs D, et al. (eds) Disorders of Hemoglobin: Genetics, Pathophysiology, Clinical Management. Cambridge University Press, Cambridge.

Hardison RC (2008) Globin genes on the move. J Biol 7:35.

Hesse JE, Nickol JM, Lieber MR, et al. (1986) Regulated gene expression in transfected primary chicken erythrocytes. Proc Natl Acad Sci U S A 83:4312–4316.

Higgs DR, Wood WG (2008) Long-range regulation of alpha globin gene expression during erythropoiesis. Curr Opin Hematol 15:176–183.

Higgs DR, Wood WG, Jarman AP, et al. (1990) A major positive regulatory region located far upstream of the human alpha-globin gene locus. Genes Dev 4:1588–1601.

Hoffmann FG, Storz JF (2007) The αD-globin gene originated via duplication of an embryonic α-like globin gene in the ancestor of tetrapod vertebrates. Mol Biol Evol 24:1982–1990.

Hofmann OM, Brittain T, Wells RM (1997) The control of oxygen affinity in the three human embryonic haemoglobins by respiration linked metabolites. Biochem Mol Biol Int 42:553–566.

Holland RA, Gooley AA (1997) Characterization of the embryonic globin chains of the marsupial Tammar wallaby, *Macropus eugenii*. Eur J Biochem 248:864–871.

Holland RA, Gooley AA, Hope RM (1998) Embryonic globins of the marsupial the Tammar Wallaby (*Macropus eugenii)*: bird like and mammal like. Clin Exp Pharmacol Physiol 25:740–744.

Hosbach HA, Wyler T, Weber R (1983) The *Xenopus laevis* globin gene family: chromosomal arrangement and gene structure. Cell 32:45–53.
Hughes JR, Cheng JF, Ventress N, et al. (2005) Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences. Proc Natl Acad Sci U S A 102:9830–9835.
Iuchi I (1973) The post-hatching transition of erythrocytes from larval to adult type in the rainbow trout, *Salmo gairdnerii irideus*. J Exp Zool 184:383–396.
Jarman AP, Wood WG, Sharpe JA, et al. (1991) Characterization of the major regulatory element upstream of the human alpha-globin gene cluster. Mol Cell Biol 11:4679–4689.
Jeffreys AJ, Wilson V, Wood D, et al. (1980) Linkage of adult alpha- and beta-globin genes in *X. laevis* and gene duplication by tetraploidization. Cell 21:555–564.
Koop BF, Goodman M (1988) Evolutionary and developmental aspects of two hemoglobin beta-chain genes (epsilon M and beta M) of opossum. Proc Natl Acad Sci U S A 85:3893–3897.
Lacy E, Hardison RC, Quon D, et al. (1979) The linkage arrangement of four rabbit beta-like globin genes. Cell 18:1273–1283.
Lauer J, Shen CK, Maniatis T (1980) The chromosomal arrangement of human alpha-like globin genes: sequence homology and alpha-globin gene deletions. Cell 20:119–130.
Maniatis T, Fritsch EF, Lauer J, et al. (1980) The molecular genetics of human hemoglobins. Annu Rev Genet 14:145–178.
Mason MM, Lee E, Westphal H, et al. (1995) Expression of the chicken beta-globin gene cluster in mice: correct developmental expression and distributed control. Mol Cell Biol 15:407–414.
Miyata M, Aoki T (1997) Head-to-head linkage of carp alpha- and beta-globin genes. Biochim Biophys Acta 1354:127–133.
Muller G, Fago A, Weber RE (2003) Water regulates oxygen binding in hagfish (*Myxine glutinosa*) hemoglobin. J Exp Biol 206:1389–1395.
Naoi Y, Chong KT, Yoshimatsu K, et al. (2001) The functional similarity and structural diversity of human and cartilaginous fish hemoglobins. J Mol Biol 307:259–270.
Niimura Y, Nei M (2005) Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. Proc Natl Acad Sci U S A 102:6039–6044.
Opazo JC, Hoffmann FG, Storz JF (2008) Genomic evidence for independent origins of beta-like globin genes in monotremes and therian mammals. Proc Natl Acad Sci U S A 105:1590–1595.
Orkin SH (1978) The duplicated human alpha globin genes lie close together in cellular DNA. Proc Natl Acad Sci U S A 75:5950–5954.
Patel VS, Cooper SJ, Deakin JE, et al. (2008) Platypus globin genes and flanking loci suggest a new insertional model for beta-globin evolution in birds and mammals. BMC Biol 6:34.
Patient RK, Elkington JA, Kay RM, et al. (1980) Internal organization of the major adult alpha- and beta-globin genes of *X. laevis*. Cell 21:565–573.
Pisano E, Cocca E, Mazzei F, et al. (2003) Mapping of alpha- and beta-globin genes on Antarctic fish chromosomes by fluorescence in-situ hybridization. Chromosome Res 11:633–640.
Proudfoot NJ, Gil A, Maniatis T (1982) The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene. Cell 31:553–563.
Proudfoot NJ, Maniatis T (1980) The structure of a human alpha-globin pseudogene and its relationship to alpha-globin gene duplication. Cell 21:537–544.
Proudfoot NJ, Shander MH, Manley JL, et al. (1980) Structure and in vitro transcription of human globin genes. Science 209:1329–1336.
Qiu Y, Maillett DH, Knapp J, et al. (2000) Lamprey hemoglobin. Structural basis of the bohr effect. J Biol Chem 275:13517–13528.
Reitman M, Felsenfeld G (1990) Developmental regulation of topoisomerase II sites and DNase I-hypersensitive sites in the chicken beta-globin locus. Mol Cell Biol 10:2774–2786.
Reitman M, Grasso JA, Blumenthal R, et al. (1993) Primary sequence, evolution, and repetitive elements of the *Gallus gallus* (chicken) beta-globin cluster. Genomics 18:616–626.
Stamatoyannopoulos G (1991) Human haemoglobin switching. Science 252:383.

Stamatoyannopoulos G (2005) Control of globin gene expression during development and erythroid differentiation. Exp Hematol 33:259–271.

Tibben EA, Holland RA, Tyndale-Biscoe CH (1991) Blood oxygen carriage in the marsupial, tammar wallaby (*Macropus eugenii*), at prenatal and neonatal stages. Respir Physiol 84:93–104.

Tyndale-Biscoe H (2005) Life Of Marsupials. CSIRO Publishing, Collingwood.

Verde C, De Rosa MC, Giordano D, et al. (2005) Structure, function and molecular adaptations of haemoglobins of the polar cartilaginous fish *Bathyraja eatonii* and *Raja hyperborea*. Biochem J 389:297–306.

Vilella AJ, Severin J, Ureta-Vidal A, et al. (2009) EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. Genome Res 19:327–335.

Villeponteau B, Martinson H (1981) Isolation and characterization of the complete chicken beta-globin gene region: frequent deletion of the adult beta-globin genes in lambda. Nucleic Acids Res 9:3731–3746.

Wagner A, Deryckere F, McMorrow T, et al. (1994) Tail-to-tail orientation of the Atlantic salmon alpha- and beta-globin genes. J Mol Evol 38:28–35.

Wells RMG (1999) Evolution of haemoglobin function: Molecular adaptations to environment. Clinical and Experimental Pharmacology and Physiology, 26:591–595.

Wheeler D, Hope R, Cooper SB, et al. (2001) An orphaned mammalian beta-globin gene of ancient evolutionary origin. Proc Natl Acad Sci U S A 98:1101–1106.

Wheeler D, Hope RM, Cooper SJ, et al. (2004) Linkage of the beta-like omega-globin gene to alpha-like globin genes in an Australian marsupial supports the chromosome duplication model for separation of globin gene clusters. J Mol Evol 58:642–652.

# CHAPTER 4: ANALYSES OF α- AND β- GLOBIN GENE LOCI IN LIZARDS

This chapter is divided into two parts; examining two different lizard species using different approaches, yet embracing a common goal of studying globin gene evolution in amniotes. The first part contains unpublished data from green anole lizard and leads to the second part that was published recently showing data from Australian dragon lizard.

## Background

As shown in Chapter 2, globin studies in the platypus showed that the bird and mammalian α-globin clusters reside in the same genomic region as do the α-β globin clusters of fish and frog. These clusters are flanked by *MPG* and *c16orf35* on 5' side and *LUC7L* on 3' end (except fish), confirming a common origin (Flint et al., 2001, Hughes et al., 2005, Hardison, 2008, Patel et al., 2008).

However, the bird and mammalian β-globin clusters reside in an olfactory region (Bulger et al., 1999, Hardison, 2005) that shares no similarities to the surroundings of their α-globin clusters. This suggests that the amniote β-globin cluster evolved by a transposition of a β-like globin gene copy from the ancient α-β globin cluster into the olfactory region, which then independently duplicated and diverged further to form the β-globin repertoire in different amniote lineages (Gillemans et al., 2003, Hardison, 2005, Hardison, 2008, Opazo et al., 2008b, Patel et al., 2008). In order to confirm this model of globin gene evolution, data on the organisation of α- and β-globin clusters is required from non-avian reptiles.

My study of the platypus α- and β-globin clusters also helped me to deduce the ancestral arrangement of these clusters in the ancestor of mammals. I showed that the ancestral mammalian α-globin cluster contained six genes in the order *HBZ-T1, HBZ-T2, HBK, HBA-T1, HBA-T2, HBA-T3*. The ancestral β-globin cluster contained either one (*HBB*) or two genes (*HBE, HBB*) according to two alternative theories for the evolution of β-globin genes in mammals. One hypothesis is that a single β-globin gene duplicated independently in the monotreme and therian lineages (Opazo et al., 2008b).

An alternate hypothesis is that it duplicated before the divergence of the mammalian subclasses, but the platypus *HBE* became homogenised by gene conversions (Patel et al., 2008).

The ancestral state is not clarified by reference to bird globin clusters. The chicken α-globin cluster, however, contains three genes (*HBP, HBK, HBA*) and the β-globin cluster contains four genes (*HBR, HBB-T1, HBB-T2, HBE*). Thus, it remains unclear as to how many genes were present in the ancestral amniote α- and β-globin clusters, what were the timings and order of their gene duplications (β-globin in particular) at the stem of amniote evolution when reptiles diverged from a line leading to mammals about 315 MYA. It would therefore be very helpful to have information on the globin clusters of non-avian reptiles.

Reptilia are an ancient group of amniotes that comprise an incredibly diverse group that accounts for more than three-quarters of all living terrestrial vertebrates, and is found in diverse regions of globe. There are three groups of Reptilia: the Archosauria (birds, crocodilians), Lepidosauria (tuatara and squamates), and Testudines (turtles). The evolutionary divergence time and phylogenetic position of these major lineages has been debated for many decades (Shedlock and Edwards, 2009 and references therein). According to Shedlock and Edwards (2009), the common ancestor of all reptiles existed 325 MYA. The first divergence of reptilians occurred in the mid-Permian period (275 Ma) to result in one line containing all archosaurs and turtles and the other line containing all lepidosaurs. The archosaurs then diverged from testudines some 231 MYA, whereas crocodilians subsequently diverged from birds around 219 MYA. In the lepidosaurian lineage, the ancient evolutionary relic, tuatara, diverged from squamates (lizards and snakes) a few years after their origin (i.e. at 272 MYA), while snakes and lizards diverged 245 MYA (Figure 1).

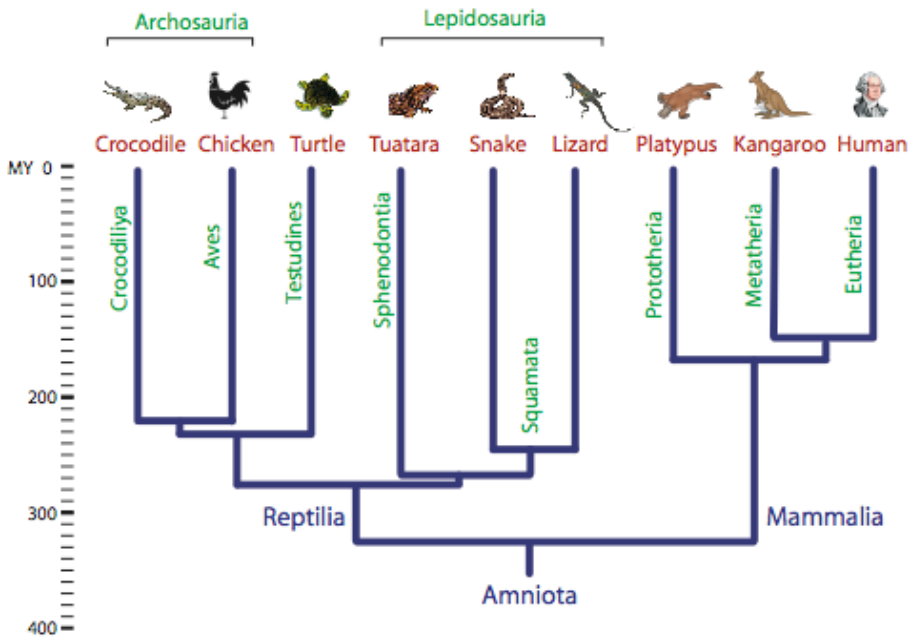



**Figure 1: Phylogenetic positions of extant reptiles**

The approximate divergence times in million years (MY) are shown for major groups of reptiles and mammals (Bininda-Emonds et al., 2007, Hedges and Kumar, 2009). The phylogenetic positions of tuatara and turtles are still uncertain.

Globin gene organization in non-avian reptiles was previously quite unknown, since research has focussed on birds and mammals. Early blood studies were limited to protein isolation and characterisation (Gorr et al., 1998 and references therein). The lack of nucleotide data for non-avian reptilian globins led to their exclusion from recent genomic studies, and since then, no detailed studies were performed on the organization of non-avian reptilian α- and β-globin clusters and their flanking genes. Therefore, it was uncertain how many genes are present in the non-avian reptilian α- and β-globin clusters, how these were arranged in the common reptilian ancestor, and how they evolved over time.

The aim of this study was to fill this important phylogenetic gap by providing data on the globins of non-avian reptiles. This would enable me to deduce the ancestral arrangement of α- and β-globin globin clusters in the reptilian ancestor, and by comparing them to those of mammals would help deduce the arrangements in different amniote lineages. These results would then enable me to study globin gene evolution in amniotes, in parallel to testing the transpositional model.

I began this project by first characterising α- and β-globin genes in the green anole using the publicly available genome sequences (Part A), and then using green anole globin sequences to isolate clones in the dragon lizard (Part B). During the preparation of a manuscript incorporating both approaches, another research group published their research on green anole globin genes (Hoffmann et al., 2010b), thereby rendering this part redundant; it was therefore removed from the main manuscript. In this thesis, I present my research from both lizard species; the green anole lizard globin genes that were removed from the published paper (Part A) and the dragon lizard globin genes that are now published (Part B).

# PART A: Characterisation of α- and β-globin genes and flanking genes in the green anole lizard

***Authors:***

**Patel, V.S.**, Ezaz, T., Deakin, J.E., and Graves, J.A.M.

***Current status of paper (circle as appropriate):***

Planned/ In Preparation/ Submitted/ Under Revision/ Accepted/ Published

***Date paper accepted for publication or anticipated date of publication:***

Excised from published paper because of prior publication of similar work. This paper will now not be published.

***Name of journal/book:***

--

***Extent to which research is your own:***

I performed all computational analyses, experiments and analysed data.

***Your contribution to writing the paper:***

I wrote the draft of the manuscript and prepared figures.

***Comments:***

This chapter (Part A) contains unpublished data from the green anole lizard, which was prepared before the appearance of a paper by Hoffmann et al. (2010b), but excised from the manuscript (leaving Part B) because it presented a similar analysis on the same data set. However, the results and conclusions of my comprehensive study into the structures of green anole globin genes differ somewhat from those of Hoffmann et al. and I also present two original models for the evolution of globin genes in reptiles. Note: Here I call *HBB1* from scaffold 3777 and *HBB2* from scaffold 7008, whereas Hoffmann et al. (2010b) refer it the other way around.

I also show in this report some experimental work that took six months of continuous BAC library screening and FISH mapping and much effort to trouble-shoot negative results. However, the results at the end were still unusable.

## Abstract

Haemoglobin gene structure and organization is well characterised in mammals and birds, but equivalent information is not available from non-avian reptiles. It is therefore not clear how many genes are contained in non-avian reptilian α- and β-globin clusters, what are their locations and expression patterns, and it is not possible to deduce the structure of ancestral reptile haemoglobin gene clusters. In this part of the manuscript, I examine the available draft genome sequences from the first squamate reptilian species to be sequenced, *Anolis carolinensis*, in order to gain insight into the organisation of reptilian globin gene clusters, their genomic context and to reconstruct the globin gene clusters at the stem of reptilian radiation. I identified two α-like globin genes in two different scaffolds, flanked by different genes. One α-like globin gene (*HBK*) had the same genomic context as α-globin genes in birds and mammals (*MPG-C16orf35-HBK-GBY*), whereas the other (*HBA*) had completely different flanking genes (*ADCY9-HBA-GSG1L*). The latter may be an artefact due to the incomplete genome assembly but may also be an indication that a transposition event has occurred in this species. Two β-like globin genes were also found on different scaffolds, suggesting a lizard-specific duplication of a β-globin gene. I propose that in the ancestor of all reptiles, only one proto-β-globin gene existed, which then duplicated in the Lepidosauria (tuatara and squamate) lineage to result in proto-$\beta^1$ and proto-$\beta^2$ as found in tuatara and snakes. In lizards, however, the progenitor of proto-$\beta^1$ was either secondarily lost, or is missing from the assembly, while the proto-$\beta^2$ gene was tandemly duplicated to result in the two β-like globin genes that I identified in the anole and those present in other lizards. There were no further duplications of the ancestral proto-β in the testudine and crocodilian lineages, but in the avian lineage it underwent independent repeated rounds of duplication. As a result, reptilian β-like globin genes form species-specific monophylic clade (except turtles) that show less homology between them and to mammals as supported by my phylogenetic analysis.

## Introduction

The recent (and still incomplete) sequencing of the genome of the first squamate species, *Anolis carolinensis* (green anole lizard) by the Broad Institute at MIT and Harvard at the depth of 6.3x coverage, could shed light into globin genes and their evolution. The first draft assembly (anoCar1.0) was released in February 2007 in the UCSC Genome browser database (http://genome.ucsc.edu). I therefore aimed to use the anole sequences to characterise α-, β-globin and their flanking genes in this species, deduce the possible arrangement of these genes and deduce the ancestral reptile globin clusters in order to study globin gene evolution. My results show the presence of *HBK* and *HBA* in two separate scaffolds but each with different flanking genes.

Two β-like globin genes (*HBB1* and *HBB2*) are also present in the anole, but their linkage with flanking genes (*RRM1, CCKBR* and *ILK*) could not be determined because all genes were located on different scaffolds in the first assembly. Although the anole genome assembly is still incomplete and lacks important data to answer our biological questions, I use existing protein data to hypothesize two models for the evolution of β-globin genes in reptiles.

## Materials and methods

### Bioinformatics and data analyses of globin genes in the green anole lizard

From the UCSC Genome Browser (http://genome.ucsc.edu), α- and β- globin genes in the *A. carolinensis* genome sequence assembly (anoCar1 - released on Feb 2007) were found on one long (~163kb) and three short scaffolds (~4.8 – 15kb). Since annotations of these genes were incomplete in the genome browser (i.e. precise location of the exons and the exact identification of the genes were not determined), these scaffolds were then analysed further.

Gene prediction programs Genscan (Burge and Karlin, 1997) and Genomescan (Yeh et al., 2001) were used to predict genes in these scaffolds using default settings. The predicted genes were then confirmed by BlastN and BlastP searches against the whole database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and also by PCR amplification from

the genomic DNA. For PCR, primers from the predicted gene sequences were designed using Primer3 v.0.4.0 (Rozen and Skaletsky, 2000) and synthesized by GeneWorks Pty Ltd (Adelaide, Aust). Details of the primers and the sizes of amplified products from gDNA are listed in Table 1.

Standard PCR reaction was performed in a final volume of 25µl, with 70ng green anole female gDNA, 1x Green GoTaq Buffer (Promega Corporation, Madison, Wisconsin, USA), 0.2mM dNTPs, 1.25U GoTaq DNA Polymerase (Promega Corporation, Madison, Wisconsin, USA) and 1µM each of forward and reverse primers. PCR cycling conditions were: 94°C for 2 mins, then 35 cycles of 94°C for 30 sec, 56-64°C for 30 sec, 72°C for 1 min, followed by a final extension of 72°C for 10 mins. The amplified products (25µl) were run on 1% TAE gel stained with SYBR® Safe (Life Technologies Corporation, California, USA). Amplified DNA from the gel was extracted using QIAquick® gel extraction kit (Qiagen Sciences, Germantown, MD, USA) following the manufacturer's protocol and were directly sequenced at the Australian Genomic Research Facility (AGRF, Brisbane, Aust). The sequences were then aligned with predicted anole globin sequences to confirm identity of the amplified products.

**Table 1:** List of primers designed from the predicted anole globin genes and the sizes of the amplified products from its gDNA

| Gene | Forward Primer | Reverse Primer | Predicted size from gDNA (bp) |
|---|---|---|---|
| *HBA* | ATGGCAAGAAGGTGGTGGA | GAGGGCGAGGATGAACTCTG | 835 |
| *HBK* | TTGGGGGAAGGTGTCTGG | CGAAGTGCGGGAAGTAGGT | 719 |
| *GBY* Exon 1-2 | AGGAGAATGGACGGCTGGT | GGCTGGAATCTGTGGGTTG | 565 |
| *GBY* Exon 3 | GGTTCTTCAGCGACTACCCTG | GGCACTTGGTGGATGTTCTT | 208 |
| *HBB1* | GGAGCAAAGTGGACATCGG | GCGTTGGAGAGGTTGCC | 345 |
| *HBB2* | CACCGTGGGGGCAGAAT | GTGGGGCTGGACAGGTTC | 190 |

### Blast search against EST library

Since tissues/RNA for anole were not available for expression studies, the predicted sequences were used to blast the *A. carolinensis* EST library (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to check for the presence of any expressed

fragments. This would have given an indication as to whether the predicted globin genes are transcriptionally active.

**Phylogenetic Analysis**

The α- and β-globin amino acid sequences of frogs and reptiles (including anole) were aligned using the program MUSCLE (Edgar, 2004), available on European Bioinformatics Institute web server (http://www.ebi.ac.uk/Tools/muscle/). The output format for the multi-alignment was chosen as 'phylip interleaved' that was then used to create phylogenetic trees using the program PhyML 3.0 (Guindon and Gascuel, 2003) with some modification to the parameters: equilibrium frequencies was set to 'empirical' and proportion of invariable sites was set to 'estimated'. Two types of phylogenetic trees were generated, one using approximate likelihood-ratio test (aLRT) method and the other using 1000 bootstrap pseudoreplicates. Both trees resulted in similar topology, so only trees generated by aLRT are shown here. Sequences from the African clawed frog (*Xenopus laevis*) were used as outgroups to root the tree. The trees were then viewed in Figtree (available for download from http://tree.bio.ed.ac.uk/stats.html).

Sequences used to create α-globin phylogeny were taken from GenBank with the following Accession numbers: indigo snake (*Drymarchon corois errabennus*) *HBA* [P0C0U6] and *HBK* [P0C0U7], sea snake (*Microcephalophis gracilis*) *HBA* [P41331], water snake (*Liophis miliaris*) *HBK* [P16417], monitor lizard (*Varanus exanthematicus albigularis*) *HBA* [P18981], common iguana (*Iguana iguana*) *HBA* [P18974], tuatara (*Sphenodon punctatus*) *HBA* [P10059] and *HBK* [P10062], Western painted turtle (*Chrysemys picta bellii*) *HBA* [P13273] and *HBK* [P02005], snake-necked turtle (*Phrynops hilarii*) *HBK* [P02006], Nile crocodile (*Crocodylus niloticus*) *HBA* [P01998], American alligator (*Alligator mississippiensis*) *HBA* [P01999], chicken (*Gallus gallus*) *HBA* [CAA42606] and *HBK* [CAA42605], frog (*X. laevis*) *HBA1* [CAA32474], and anole lizard *HBA* and *HBK* [from this study].

Sequences used to create β-globin phylogeny were: indigo snake *HBB1* [Accession number: P0C0U8], common iguana *HBB1* [P18987] and *HBB2* [P86390], sea snake *HBB2* [P41332], water snake *HBB1* [P16418], Indian cobra *HBB2* [P22743], Indian python (*Python molurus bivittatus*) *HBB1* and *HBB2* (Stoeckelhuber, 1992, Gorr et al.,

1998) monitor lizard *HBB1* [P18993], marine iguana *HBB2* [P86391], tuatara *HBB1* ($\beta^A$) [P10060] and *HBB2* ($\beta^D$) [P10061], western painted turtle *HBB* [P13274], loggerhead sea turtle *HBB* [Q10733], red-footed tortoise (*Geochelone carbonaria*) *HBB* [Q98905], Nile crocodile *HBB* [P02129], American alligator *HBB* [P02130], spectacled caiman *HBB* [P02131], chicken *HBR* [AAD03345], *HBBa* ($\beta^A$) [AAD03347] and *HBBh* ($\beta^H$) [AAD03346], Zebrafinch *HBR, HBBa* ($\beta^A$)*, HBBh* ($\beta^H$) (Alev et al, 2009; supplementary data), duck (*Stictonetta naevosa*) *HBBa* [ACT81585], goose (*Branta canadensis*) *HBB* [P68944], swan (*Cygnus olor*) *HBB* [P68945], macaw (*Ara ararauna*) *HBB* [P02116], cormorant (*Phalacrocorax carbo*) *HBB* [P10782], parakeet (*Psittacula krameri*) *HBB* [P21668], frog adult *HBB2* [P02132], larval *HBB1* [P02137] and larval *HBB2* [P02133], and anole lizard *HBB1* and *HBB2* [from this study].

**BAC library screening**

Three or four sets of overgos for each predicted globin gene including their flanking genes were designed using OvergoMaker (Washington University Genome Sequencing Centre 2010). Table 2 lists 31 overgos used in this experiment. These overgos were labelled with radioactive isotope $^{32}$P-dATP and $^{32}$P-dCTP, and then hybridised to high-density filters from green anole genomic bacterial artificial chromosome (BAC) library (CHORI-318 from Children's Hospital Oakland Research Institute, USA) according to the protocol as described in Ezaz et al. (2009). The presence of globin and/or flanking genes in the BAC clones was confirmed by dot blots, following protocol described in Patel et al. (2008).

**Mapping by fluorescence *in situ* hybridisation (FISH)**

Chromosome preparations of the male and female green anole lizards were kindly provided by Dr Dan Janes (Harvard University, USA). The BAC clones containing globin and/or flanking genes were labelled with fluorchrome-labelled dUTPs (SpectrumGreen and SpectrumOrange) and hybridised to green anole chromosomes using either single- or two-colour FISH, following protocols described in Deakin et al. (2008) and Ezaz et al. (2009).

**Table 2:** List of overgos used in this experiment

| GENES | FORWARD OVERGO | REVERSE OVERGO |
|---|---|---|
| *HBA* | GACCTCCATGCCGAGAAGCTCCGT | AGTTGACAGGGTCAACACGGAGCT |
| | TGAGGACGACAAGAACCACATCCG | ATGGCCCCAGATGGCCCGGATGTG |
| | TGCTGACCGCTGAGGACCGCAAAC | CAAATGGCTGTCAGCAGTTTGCGG |
| | CTCCAAGGACATCCAAGTGCATGG | TCTGGCCACCTTCTGGCCATGCAC |
| *HBK* | GCTGATCAAGTGCTTCCACGTGGT | CAGGTGAGTGGCCAGAACCACGTG |
| | ACCTGCATGCCTACAACCTCCGCG | AAGTTGACTGGGTCCACGCGGAGG |
| | CACCACACCACCAAGACCTACTTC | TCATGTCATAGTGCGGGAAGTAGG |
| | CCTTGAAGCCAGAGTTCATCCTCG | AAACACTTGTCGAGGGCGAGGATG |
| *GBY* | CCTTCAGCCAAGTGATTGAGAACA | TGGTTCCAGTTCTCCATGTTCTCA |
| | GCGGCACATCCGGGAAATCTGGAC | ATTTTCAAAGGCCGCGGTCCAGAT |
| | TTCAGCGACTACCCTGCCAGCAAG | CCGTCTTGAAGTACTGCTTGCTGG |
| *HBB1* | TGGACCCAGAGGTTCTTCCCCGAC | TGGAGAGGTTGCCAAAGTCGGGGA |
| *HBB2* | CCGCCTACCAGAAGCTGGTCAACG | AAGGCGTGGGACACCACGTTGACC |
| *HBB1/HBB2* | GGACAACATCAAGGACACCTTCGC | CAGCTCGCTCAGCTTGGCGAAGGT |
| *MPG* | GAGAGGGGGCAGCGGTGCTGCTGC | AGGGGCTCTAGGGAGCGCAGCAGC |
| | TATCTCCGGGGCAACAAGTACGTC | CCTTGTCAACAACGCTGACGTACT |
| *C16orf35* | AGTGATGACATGACGCTCACCAGC | AGTTGTCCATGCTGGGGCTGGTGA |
| | TCAGCGTTATCCTGGTCAGCTCTG | TTGTTCCCGCGGCTGCCAGAGCTG |
| | AATTCTGGCCACCAAGTCGGACAT | AAACTTCTTGCCGCACATGTCCGA |
| *LUC7L* | AAACGGTAGCAGAAAAACAGGAGA | CGGTCCTGATTTCTCTTCTCCTGT |
| | TTCACAGATGAGCGAGTCTGCAAG | AATCCAGAAGGTGGCTCTTGCAGA |
| | AGAGCACCAACGGCAAATCTCGCT | TCCTCCGACCTCTTCGAGCGAGAT |
| *RRM1* | AAGATCACCTCCCGGATCCAGAAG | TAAGACCGTAGCACAGCTTCTGGA |
| | TGAAAGTGATCCAGGGCCTCTACA | ACAGTGGTGACACCACTGTAGAGG |
| | ATTGACCAGAGCCAGTCCCTGAAC | GCTCAGCAATGTGGATGTTCAGGG |
| *ILK* | ATCCCCTACAAGGACACCTTCTGG | TCCGAGTGGTGCCCTTCCAGAAGG |
| | TGGTCAGCATCTGCAACAAGTACG | TCCAGCGGGGTCTCTCCGTACTTG |
| | GACATGACGGCCCGGATCAGCATG | AGAACTTGACGTCGGCCATGCTGA |
| *CCKBR* | TCACCAACTCCTTCCTGCTCTCGC | AGGTCGCTGAGGGCCAGCGAGAGC |
| | ATCAACCCCTTCGTCTACTGCTTC | GGAAGCGCTTGTTCATGAAGCAGT |
| | TGGTGGCCATCGCCATTGAGCGCT | TTGCAGATGGCGTTGTAGCGCTCA |

## Results

### Characterisation of alpha and beta-globin gene scaffolds in the green

Three years after the first publication of the draft genome assembly of the green anole lizard, *A. carolinensis* (anoCar1 - Feb 2007) on the UCSC Genome Browser (http://genome.ucsc.edu), its current assembly is still incomplete for α- and β-globin clusters, as individual genes appear on different short scaffolds (Table 3). Even the annotations (such as exact locations of the exons, positions of splice sites, cDNA sequences, the lengths of the introns and exons, and the chromosomal location) of the globin genes are currently incomplete. Therefore, these scaffolds were analysed further to gain insight into the structure and gene composition of the lizard α- and β-globin clusters.

**Table 3:** Scaffold identities, their sizes and annotated genes on the UCSC Genome Browser for *A. carolinensis* (anoCar1)

| Scaffold ID | Scaffold size (bp) | Genes annotated on it by UCSC |
|---|---|---|
| 2790 | 15,038 | *C16orf35*, α-related, *MPG* |
| 1188 | 163,036 | α-related (Position 107,470 - 108,622) |
| 3777 | 10,310 | β-related |
| 7008 | 4,809 | β-related |
| 985 | 271,130 | *RRM1* (Position 66,082 - 113,876) |
| 1207 | 168,833 | *CCKBR* (Position 4,044 - 9,871) |
| 3689 | 10,604 | *ILK* (Position 1,058 - 7,987) |

Using the sequence information from four scaffolds (ID: 1188, 2790, 3777 and 7008), the genes contained in them were predicted by Genscan and Genomescan. The predicted genes were then identified by BlastN and BlastP searches against the nr database and also confirmed by PCR amplification from anole gDNA. The identities of the predicted genes were also confirmed by phylogenetic analysis.

From the above four scaffolds, a total of five globin genes were predicted, which had characteristics of all vertebrate α- and β-globin genes such as a three-exon/two-intron structure and conserved splice sites (Table 4).

**Table 4:** (A) Locations of the predicted globin genes in the *anoCar1* Genome Browser and (B) the sizes of their respective exons, introns, coding domains and amino acid polypeptide chains

(A).

| Scaffold ID | Gene identified | Exon 1 | Exon 2 | Exon 3 |
|---|---|---|---|---|
| 1188 | *HBA* | 107470 – 107561 | 107660 – 107864 | 108497 – 108625 |
| 2790 | *HBK* | 4064 – 3970 | 3351 – 3147 | 3048 – 2920 |
| 2790 | *GBY* | 697 – 785 | 1984 – 2206 | 2307 – 2438 |
| 3777 | *HBB1* | 4517 – 4608 | 4845 – 5067 | 7670 – 7798 |
| 7008 | *HBB2* | 1005 – 1096 | 1194 – 1416 | 2957 – 3085 |

(B).

| Genes | Exon 1 (bp) | Intron 1 (bp) | Exon 2 (bp) | Intron 2 (bp) | Exon 3 (bp) | CDS (bp) | Polypeptide (aa) |
|---|---|---|---|---|---|---|---|
| ***HBA*** | 92 | 100 | 205 | 634 | 129 | 426 | 141 |
| ***HBK*** | 95 | 620 | 205 | 100 | 129 | 429 | 142 |
| ***HBB1*** | 92 | 238 | 223 | 2604 | 129 | 444 | 147 |
| ***HBB2*** | 92 | 99 | 223 | 1542 | 129 | 444 | 147 |
| ***GBY*** | 89 | 1200 | 223 | 102 | 132 | 444 | 147 |

## Alpha-globin genes in the green anole

Two of the predicted genes were identified as α-like; one in scaffold 1188, and the other in scaffold 2790 (Tables 3 and 4). These two genes were quite different from each other, showing a 46% identity score at the protein level, and they contained slightly different polypeptide lengths. The predicted polypeptide in scaffold 1188 encoded 141 aa, which obtained the highest identity score with *HBA* ($\alpha^A$-globin) of other reptiles and birds; of 74% with the common iguana (*Iguana iguana*), 66% with Texas indigo snake (*Drymarchon corais erebennus*) and 55% with chicken (*Gallus gallus*). Phylogenetic analysis of all reptilian α-globin amino acid sequences showed that this gene grouped with other squamate *HBA* with high support of 0.986, and with other reptilian *HBA* with 0.567 support (Figure 1). Hence, this gene was named *HBA,* based on its close relationship with other reptilian *HBA*.

The other predicted polypeptide in scaffold 2790 encoded 142 aa, which obtained the highest protein identity score with *HBK* ($\alpha^D$-globin) of snakes and birds; 66% with

monocled cobra (*Naja kaouthia*) and Japanese rat snake (*Elaphe climacophora*), and 60% with chicken *HBK*. Phylogenetic analysis grouped this gene with other squamate and reptilian *HBK* with high ratios of 0.994 and 0.959 respectively (Figure 1). Hence, this gene was named *HBK*. However, exon 2 (193 bp) of the predicted sequence was found to be shorter than other α-like globin genes (205 bp). Sequence alignment of this gene with other reptilian *HBK* genes indicated that it lacked 12 bp encoding 4 aa (VNFK) at the end of its exon 2. Upon further investigation, I found that Genscan and Genomescan predicted all genes based on the splice site of GT/AG. However, the splice site of exons 2 and 3 of all reptilian *HBK* genes studied to date, is somewhat different from that of other α-globin genes, containing a GC/AG rather than a GT/AG splice site. This unique characteristic was also present in the putative anole *HBK*, so the length of exon 2 was corrected to 205 bp. The extra 12 bp encoded VNFK and the presence of GC splice site at the end of exon 2 are well conserved with all other reptilian *HBK* genes.

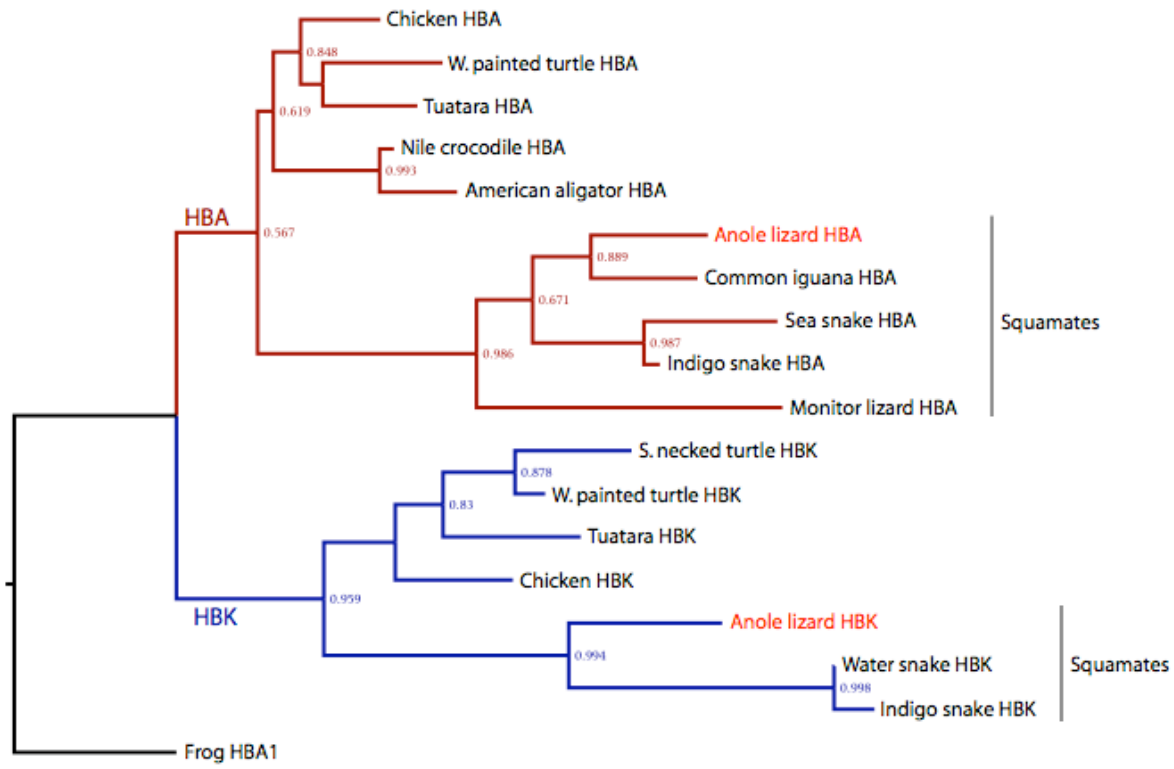



**Figure 1: Evolutionary relationship among reptilian α-like globin proteins using frog as an outgroup.** Protein sequences of reptilian and frog α-like globin genes were used to generate phylogenetic tree in PhyML using aLRT method. Accession numbers for the sequences are shown in methods and materials. Numbers adjacent to branches refer to approximate likelihood ratio.

The predicted DNA sequences of both α-like globin genes were used to BLAST search the anole EST library to gain insight into their expression. Out of both globins, only *HBK* obtained 100% hits with two identical mRNA sequences isolated from a dewlap cDNA library [Genbank EST Accession numbers: FG672953 and FG673001]. It also obtained partial hits with two mRNA sequences from whole brain cDNA library [FG662920 and FG662968] and one mRNA sequence from a mixed (tongue, liver, gallbladder, spleen, heart, kidney and lung) cDNA library [FG732129]. This suggests that *HBK* is transcriptionally active in anole, with confirmed expression in brain and dewlap (the organs unexpected for globin gene expression or synthesis of red blood cells). It may also be expressed in spleen, which is a minor site for erythropoiesis in some reptiles (Dabrowski et al., 2007 and references therein), but because the mRNA sequence was retrieved from a mixed sample of tissues, this could not be confirmed. Unfortunately, no hits were obtained for *HBA*, suggesting that it is not expressed in tissues that were sequenced (i.e. dewlap, brain, tongue, liver, gallbladder, spleen, heart, kidney and lung) leaving open the possibility that it is expressed in some other tissues or at different developmental stages.

The number of genes present in the anole is different from the gene composition and organization of globins in birds, in which there are three genes in the α-globin cluster (*HBP, HBK, HBA*). I searched for the third gene (*HBP*), which is the homologue of an early-expressed α-globin gene with expression in the embryos of birds and other mammals but I could not find any trace of it in the anole genome browser.

**Beta-globin genes in the anole lizard**

Two genes were predicted to be β-related; one in scaffold 3777 and the other in scaffold 7008 (Tables 3 and 4). These two genes were quite similar to each other, having identity scores of 86% at the nucleotide level and 75% at the protein level. The predicted gene in scaffold 3777 encoded a polypeptide of 147 aa, which obtained high protein identity scores of 83% with iguana *HBB1*, 73% with monitor lizard (*Varanus exanthematicus albigularis*) *HBB1* and 71% with chicken *HBE*. The phylogenetic analysis of all reptilian β-globin amino acid sequences showed that this gene grouped with common iguana with a high support of 0.982 although support with other lizard *HBB1* was low (Figure 2). Hence, this gene was named *HBB1* based on its close relationship with other lizard *HBB1*.

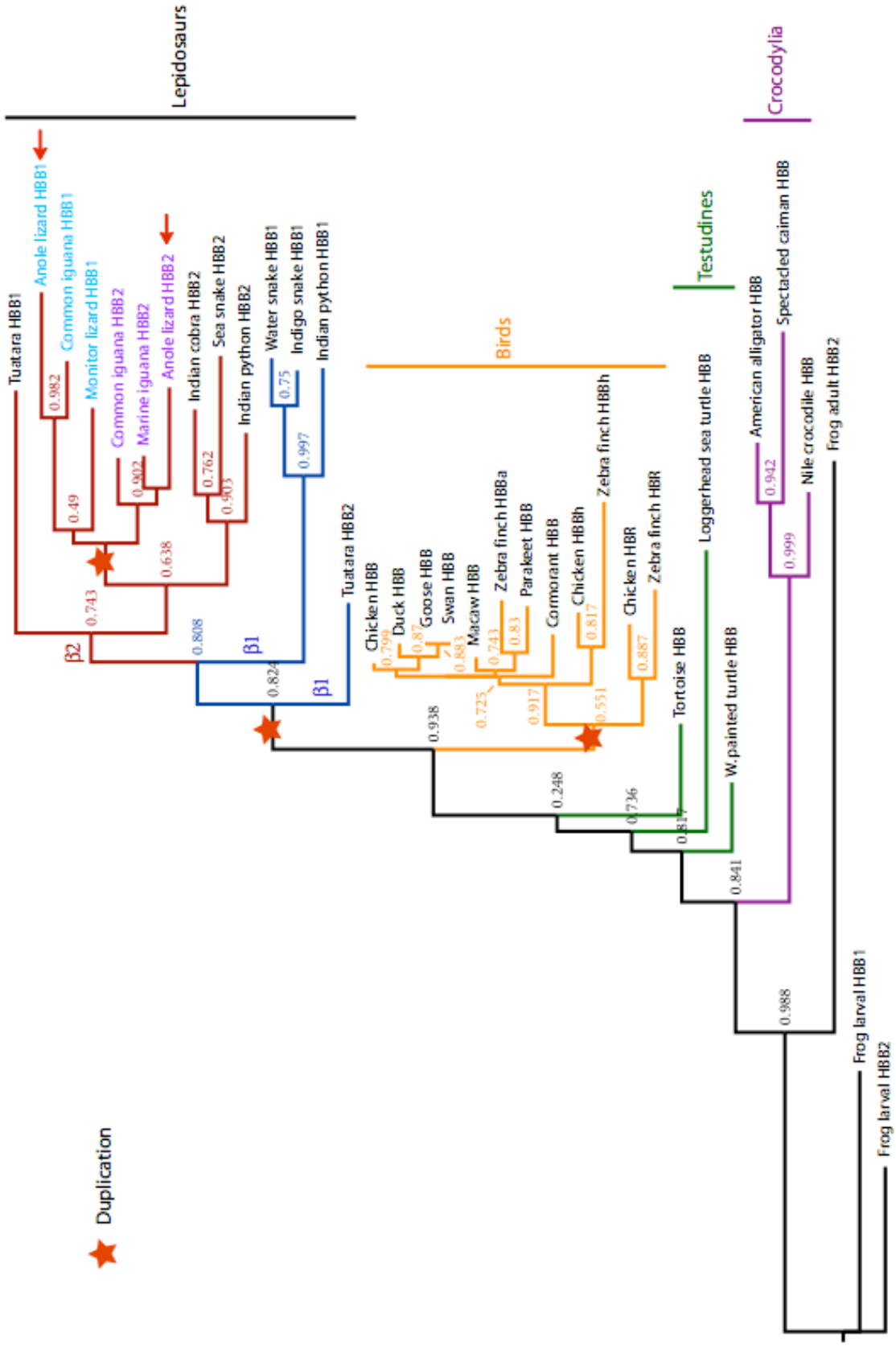

Lepidosaurs
Crocodylia
Testudines
Birds
Duplication
Tuatara HBB1
Anole lizard HBB1
Common iguana HBB1
Monitor lizard HBB1
Common iguana HBB2
Marine iguana HBB2
Anole lizard HBB2
Indian cobra HBB2
Sea snake HBB2
Indian python HBB2
Water snake HBB1
Indigo snake HBB1
Indian python HBB1
Tuatara HBB2
Chicken HBB
Duck HBB
Goose HBB
Swan HBB
Macaw HBB
Zebra finch HBBa
Parakeet HBB
Cormorant HBB
Chicken HBBh
Zebra finch HBBh
Chicken HBR
Zebra finch HBR
Tortoise HBB
Loggerhead sea turtle HBB
W.painted turtle HBB
American alligator HBB
Spectacled caiman HBB
Nile crocodile HBB
Frog adult HBB2
Frog larval HBB1
Frog larval HBB2
β2
β1
β1
0.982
0.49
0.743
0.902
0.638
0.762
0.903
0.808
0.75
0.997
0.824
0.799
0.87
0.883
0.725
0.743
0.83
0.917
0.817
0.551
0.938
0.887
0.248
0.736
0.942
0.999
0.841
0.988

**Figure 2 (previous page): Evolutionary relationship among reptilian β-like globin proteins using frogs as outgroups.** Protein sequences of reptilian and frog β-like globin genes were used to generate phylogenetic tree in PhyML using aLRT method. The tree highly supports species-specific clustering of β-like globin genes in reptiles. The presence of multiple genes in birds and lepidosaurs indicate independent duplication of an ancestral β (denoted by *red* star) in these lineages and in lizard lineage. Accession numbers for the sequences are shown in methods and materials. Numbers adjacent to branches refer to maximum likelihood ratio.

The other predicted gene in scaffold 7008 also encoded a polypeptide of 147 aa, which obtained high protein identity scores of 78% with monitor lizard *HBB1*, 77% with chicken *HBB* and 76% with Indian cobra snake (*Naja naja*) *HBB2*. The phylogenetic analysis grouped this gene with other lizard *HBB2* with 0.902 support (Figure 2). Hence, this gene was named *HBB2*.

The phylogenetic analyses (Figure 2) presented here for all β-globin proteins in reptiles revealed two interesting relationships. Firstly, it grouped all lizard *HBB1* and *HBB2* as a sister clade to snake *HBB2* with 0.638 support, suggesting a lizard-specific duplication of a snake-like *HBB2* gene. Snakes and lizards both possess two β-like globin genes that were thought to be a result of squamate-specific duplication of an ancestral β-globin gene (Gorr et al 1993), therefore, I expected that lizard *HBB1* would group with snake *HBB1* and likewise lizard *HBB2* with snake *HBB2*. However, this was not the case here, as both aLRT and bootstrap supported phylogenies showed snake *HBB1* forming a monophylic clade separate to those containing snake *HBB2*, tuatara *HBB1*, lizards *HBB1* and *HBB2* with a high support of 0.808. The other notable feature of the phylogenetic tree was the grouping of all β-like globin genes in a species-specific manner (except turtles where signals are lost) rather than gene-specific manner (as in the case of α-globins). For example, all bird β-like globin genes formed a monophylic clade distinct to those of lepidosaurs (tuatara and squamates).

Moreover, the predicted sequences of both *HBB1* and *HBB2* were used to BLAST search the anole EST library to check for their expressions. There were no hits to either gene, suggesting that they are expressed in tissues other than those sequenced. The entire anole genome database was also searched for embryonic β-like globin genes

using chicken *HBE* and *HBR* and mammalian *HBE* genes, but no orthologue of these genes were present in the genome browser.

**Discovery of the first reptilian Globin Y (*GBY*)**

I discovered a fifth globin gene that was not annotated on the anole genome browser. This gene was found in scaffold 2790, approximately 600 bp from *HBK*. It bears homology to *GBY*, a gene recently discovered in amphibians (Fuchs et al., 2006) and platypus (Patel et al., 2008). Hence, I named this gene *GBY*. Again, Genscan and Genomescan could not predict exon 1 correctly, lacking the start site and splice site of GC rather then GT. Upon further comparisons with platypus and frog *GBY* sequences, I was able to show that exon 1 of the anole *GBY* contains a start codon (ATG) at position 697 of scaffold 2790, but is different in length (89 bp) compared to exon 1 of the frog and platypus *GBY* (98 bp). Moreover, like anole *HBK*, *GBY* also has a splice site of GC/AG rather than GT/AG between exons 1 and 2. As a result, like other α, β and *GBY* globins, the predicted anole *GBY* also has a three-exon/two-intron structure and conserved donor/acceptor (GT/AG and/or GC/AG) (Table 4). The lengths of its exons 1, 2, and 3 were 89 bp, 223 bp, and 132 bp respectively, compared with 98 bp, 223 bp and 144 bp in platypus *GBY* and 98 bp, 232 bp and 141 bp in *X. laevis GBY*. Because of its shorter exons 1 and 3, the predicted polypeptide of anole *GBY* (147 aa) was shorter than the platypus *GBY* (154 aa) and frog *GBY* (156 aa). Its length is the same as other β-like globin genes, suggesting that the ancestral gene might have been of this length. Moreover, the anole *GBY* gene contained the shortest introns; 1200 bp (intron 1) and 102 bp (intron 2) compared to *GBY* from other species that are usually more than 3 kb and can be as long as 21 kb (intron 1 of *X. laevis GBY*).

Unlike the α- and β-like globin genes, the nucleotide and amino acid sequences of *GBY* are not highly conserved among different species. In this case, at the protein level the anole *GBY* showed an identity score of 36% with platypus *GBY* and 38% with *X. laevis GBY*, which is approximately the same as the alignment score between platypus and frog *GBY* (Patel et al., 2008).

**Genes flanking the anole α- and β-globin genes**

In scaffold 2790, two other genes were present adjacent to the anole *HBK* and *GBY* genes that were identified as orthologues of *C16orf35* and *MPG*. These four genes were arranged in the order 5'-*MPG-c16orf35-HBK-GBY*-3', the order that has been conserved in all jawed vertebrates.

However, the genes flanking the anole *HBA* on scaffold 1188 were totally different to those flanking all known globin genes. The anole *HBA* was flanked by *ADCY9* on one side and *GSG1L* on the other. In human these genes also reside on chromosome 16, but lie further downstream from the human α-globin cluster. Upon further investigation of this scaffold, I found a gap located 3' of the anole *HBA* gene indicating that it might have been incorrectly assembled in the genome browser.

The scaffolds that contain *HBB1* and *HBB2* were extremely short and did not contain any additional genes. *RRM1, CCKBR* and *ILK*, genes which flank the β-globin cluster in other amniotes (Taylor et al., 2006, Hardison, 2008, Patel et al., 2008), were found on different scaffolds; *RRM1* in scaffold 985, *ILK* in scaffold 3689, and *CCKBR* in scaffold 1207. Since it was not clear from the anole genome browser if these genes flank the anole β-like globin genes, I attempted to isolate BAC clones containing these genes and map their locations on anole chromosomes using FISH.

**Screening anole BAC library for globin and flanking genes in order to map their locations**

To determine whether (i) anole *HBA* and *HBK* are located together, (ii) *RRM1*, *CCKBR* and *ILK* flank the β-globin cluster, and (iii) whether α- and β-like globin genes are located together on one chromosome or separated on different chromosomes, I screened the anole BAC library with 31 overgos (each gene had three to four pairs of overgos to increase the success of screening) several times to isolate clones containing α, β and flanking genes. As a consequence, 15 clones were identified but only two (CH318-100L29 and CH318-222G23) were found positive for both *HBB1* and *HBB2*, and none for α-like globin or other genes. Multiple attempts to map both clones onto the male and female anole chromosomes using single- and two-colour FISH failed to reveal their locations. These experiments either resulted in no fluorescent signals being detected or the signals were not clear and consistent on the chromosomes. This appeared to be

largely a consequence of poor quality metaphases preparations, with metaphase chromosomes of male anole being hardly visible and the non-specific hybridisation on female chromosomes was so great that it was almost impossible to locate true signals from background.

## Discussion

Genome sequencing projects have permitted research to an unprecedented depth and breadth, providing new opportunities to understand genome organization, function and evolution. Studying the gene content and flanking regions of the α- and β-globin clusters in mammals and other vertebrates should be greatly facilitated by these sequencing projects. In practice, however, many draft genomes have not been fully assembled to the extent that all genes are annotated and contigged together. Many gaps interrupt the sequences, and genes of interest lie on different contigs, or may be missing from the assembly altogether. Thus, there is a huge and continuing effort to overcome these problems and fill in the missing pieces to the puzzle.

With the availability of the draft version of the green anole genome sequences at the depth of 6.3 times or more, I was able to answer some questions about the structure and organization of the lizard globin clusters. However, the anole data set is still incomplete and rigorous analysis must await completion of this project.

### Discovery and significance of *HBK* and *HBA* in lizards and other reptiles

The anole genome contains at least two α-like globin genes identified as orthologues of *HBK* and *HBA,* with a confirmed expression of *HBK* in dewlap and brain. Both *HBK* and *HBA* are transcriptionally active and form functional haemoglobin in other reptiles (Hiebl et al., 1987, Hiebl et al., 1988, Rucknagel and Braunitzer, 1988, Fushitani et al., 1996, Melo et al., 2003). Phylogenetic analysis of reptile *HBA* and *HBK* presented in this study is consistent with that of Gorr et al. (1998). The orthologues of both genes are present in all mammals, suggesting that these two genes were present in the common ancestor of all amniotes.

However, the function of *HBK* in amniotes is speculative. In both reptiles and mammals, the *HBK* is expressed in all stages of development (embryonic and adult), but in mammals its product does not assemble into functional haemoglobin tetramers (Goh et al., 2005, Cooper et al., 2006, Patel et al., 2008), whist in reptiles it does. The reptilian *HBK* has a higher affinity for oxygen than *HBA* and therefore, it has been suggested to confer an advantage to the embryo living in the hypoxic environment of an egg, and to the adult surviving in a hypoxic condition related to high-altitude flight or prolonged submersion (Goh et al., 2005). The expression of *HBK* in an adult anole suggests that this gene may also be functional in the anole, but its expression in tissues that are not involved in erythropoiesis or globin synthesis suggests that it might have a special function in addition to oxygen transportation to adapt to the species's lifestyle.

**Is there an embryonic α-like globin in non-avian reptiles?**

After the duplication of a single primordial globin into the ancestral α- and β-globin genes (~450 MYA), the ancestral α-globin gene is thought to have first duplicated into progenitors of early- (*HBP*) and late- (*HBA*) expressed globin genes (~400 MYA), followed by another tandem duplication, to result in the *HBK* lineage in the common ancestor of all tetrapods (Czelusniak et al., 1982, Goodman et al., 1987, Cooper et al., 2006, Hoffmann and Storz, 2007). Therefore, these three gene lineages (*HBP, HBK, HBA*) have been duplicating and diversifying for ~400 MYA in different tetrapod lineages.

However, we were unable to find any sign of an embryonic *HBP* in the anole assembly. Given that birds contain an early-expressed *HBP*, its absence in the anole genome browser was unexpected. It is possible that *HBP* is present in the anole genome but was missed from the incomplete genome sequencing and assembly. Alternatively, it is not present in their genome. This would explain early studies at the protein level which separated distinct embryonic and adult haemoglobins by electrophoresis; however, these molecules did not contain *HBP* or any other embryonic β-like globin chains (McCutcheon, 1947, Manwell, 1960, Pough, 1969, Pough, 1977, Isaacks et al., 1978).

To our knowledge, no embryonic globin genes have ever been characterised in any non-avian reptiles to date, so it is possible that genomes of non-avian reptiles lack embryonic α-globin genes. This would imply that *HBP* was secondarily lost from the genomes of only non-avian reptiles including the anole, as this gene is present in the

other amniotes including chickens and mammals, so it must have been present in the genome of the common amniote ancestor (Figure 3). Its absence would then lead us to speculate that the anole has recruited either *HBK* or *HBA* gene to fulfil its embryonic requirement of oxygen. Obviously, further expression studies are needed to test whether *HBK* or *HBA* is expressed in the early stages of development.

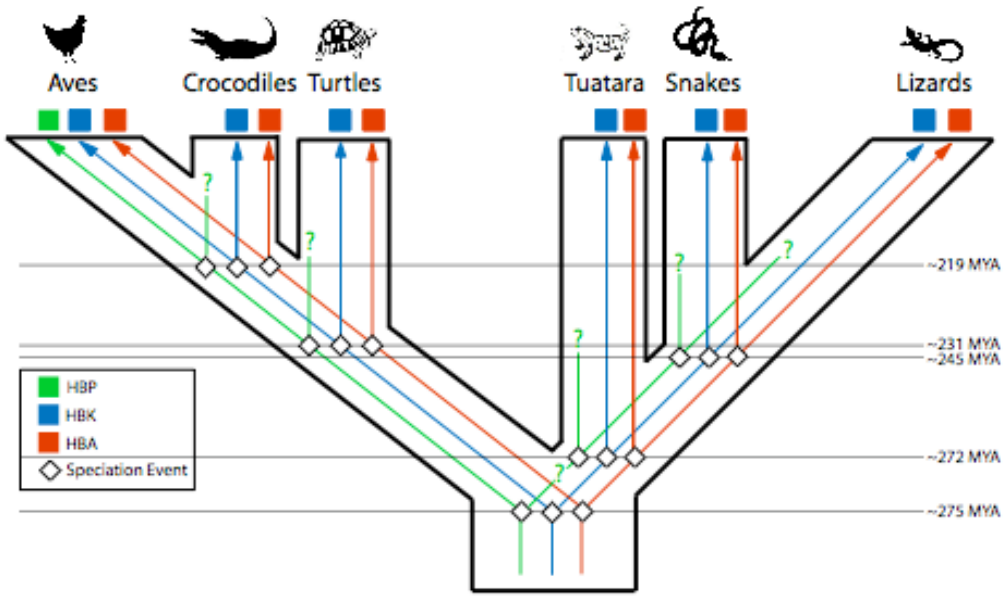



**Figure 3: Model of the evolution of α-globin genes in reptiles**

According to this model, the reptilian ancestor already contained three α-like globin genes, *HBP* (left branch in *green*), *HBK* (middle branch in *blue*) and *HBA* (right branch in *red*) of which, *HBK* and *HBA* have been present in all extant reptiles. *HBP* is found in birds only but it is uncertain if it has been secondarily from genomes of all non-avian reptiles (by three or more independent events leading to its pseudogenisation or total elimination) or its presence has been undetected so far.

### Discovery of *GBY* in lizards and possible roles

The discovery of *GBY*, a more distantly related member of globin super-family, adjacent to *HBK* gene in the anole, sparked renewed interest into its evolutionary origin and function. Its orthologue was first discovered in frogs (Fuchs et al., 2006) and then in platypus (Patel et al., 2008). It has been recently shown that *GBY* forms a small monophylic clade, sharing a common ancestor with other vertebrate-specific globins (myoglobin, cytoglobin, Globin E) (Hoffmann et al., 2010a). Although the *GBY* in frog, platypus and lizard is a neighbour to the α-globin cluster, it does not interact with the

members of the cluster. Its function and importance is currently unknown, but studies in frogs have revealed that *GBY* bears all the hallmarks of a functional respiratory protein, suggesting that it has some sort of oxygen-related function (Fuchs et al., 2006).

Expression studies of *GBY* could not be performed in the anole, due to unavailability of tissue and RNA from this species. However, expression studies of *GBY* have previously been performed in frogs (Fuchs et al., 2006) and platypus (Patel et al., 2008). In both species, it is expressed in broad range of tissues but most strongly in gonads, unlike any other globin gene. The significance of this unusual expression is still not well understood, but it is believed that *GBY* may be involved in some sort of special oxygen transportation in gonads of these species (Patel et al., 2008).

Both these species and lizards have one major phenotype in common, that is, they are all oviparous (lay eggs). Frogs lay slimy and gel-like eggs in puddles, ponds and lakes, lizards lay soft and leathery eggs in moist soil/sand, and platypuses lay soft and leathery eggs in burrowed soil. Perhaps the hypoxic conditions under which eggs are incubated demand a special oxygen transportation system that enables the embryos to survive. If this were the case, we might expect a functional *GBY* to be present in the genomes of oviparous, but not viviparous amniotes. However, I could find no *GBY* in chicken and zebrafinch genomes, although they also lay eggs. *GBY* could therefore, have been lost secondarily in birds after they acquired flight, a different structured eggshell with more pores, a drier environment for egg incubation, and more nurturing from their mother. It would be interesting to search for *GBY* in other birds, and also in fish.

Comparative analyses of *GBY* between different species show considerable interspecies variation in sequence and structure. The nucleotide sequence of *GBY* is poorly conserved (<50%) among different species; for instance *X. laevis GBY* shows only 38% amino acid identity with both platypus and lizard *GBY*, and lizard and platypus *GBY* show only 36% amino acid identity. This is, perhaps not surprising, given the very long divergence times of these species. However, even closely related species show considerable divergence; for example, *X. tropicalis* and *X. laevis GBY* have only 79% amino acid identity (Fuchs et al., 2006). Even the lengths of the coding domain of *GBY* differ between frog (156aa), lizard (147aa) and platypus (154aa). The intron lengths, in particular, are very different between these species; for example, the introns of *GBY* in frogs are enormous compared to *GBY* introns in lizards.

These differences in *GBY* sequence and structure could be a result of the ancient origin and rapid evolution of *GBY*. Since *GBY* was present in the common ancestor of all jawed vertebrates (Patel et al., 2008, Hoffmann et al., 2010a), it has been diversifying for more than 400 million years, and evolving at a much higher rate compared to other globin genes. For example, in closely related species such as *X. laevis* and *X. tropicalis* that diverged about 30 MYA, *GBY* shows a higher amino acid replacement rate of $1.29x10^{-9}$ per site per year, compared to $0.54x10^{-9}$ per site per year for mammalian and *Xenopus* cytoglobin based on a divergence time of 360 MYA (Fuchs et al., 2006).

So far, no orthologue of *GBY* has been discovered in any therian (marsupial and eutherian) mammal, suggesting that it was lost before the divergence of marsupials from eutherians 148 MYA. Perhaps its rapid evolution made *GBY* more vulnerable to nucleotide changes that led to its pseudogenisation and loss in therian mammals, where its function may have become redundant with the evolution of viviparity.

**Findings and significance of *HBB1* and *HBB2* in lizards**

Two birds (the chicken and zebrafinch) are the only species of reptiles for which the entire globin clusters have been evaluated at the molecular level. Early studies of the globins in other reptiles relied on protein data that showed the presence of certain β-globin chain in the blood of either an embryo (that combined with *HBK*) or an adult (that combined with *HBA*), implying that genes encoding these proteins are also present in their genomes. The relationship between the reptile β-globin chains was last analysed more than a decade ago (Gorr et al., 1998).

Birds possess four β-like globin genes (*HBE, HBB-T1, HBB-T2, HBR*). However, so far in non-avian reptiles the score has been spotty; one β-globin (*HBB*) in turtles and crocodiles, and two in squamates (*HBB1* and *HBB2*) and tuatara (*HBB1* binds with *HBA* and *HBB2* binds with both *HBA* and *HBK*) (Gorr et al., 1998 and references therein). It is possible that non-avian reptiles possess more β-globin like genes, but the incomplete sequencing and assembly have not discovered them.

I discovered that the anole genome contains two paralogous β-like globin genes. These were identified as *HBB1* and *HBB2*, which had not been previously identified by protein

analysis in this species. Phylogenetic analysis of reptile β-globin protein data presented in this report is inconsistent with that of Gorr et al. (1998). Using very limited β-globin protein sequences, Gorr and colleagues reported that lizard *HBB1* and *HBB2* were closely related to snake *HBB1* and *HBB2*, respectively, suggesting a squamate-specific duplication of an ancestral β-globin gene. However, their phylogenetic analysis lacked tuatara *HBB1* and many other bird β-globin sequences including chicken *HBE*, *HBP* and *HBB-T1*. My phylogenetic analysis contains more β-globin protein sequences of extant reptiles than used by Gorr et al. (1998) so is likely to be more robust.

My phylogenetic analysis showed that the lizard (including anole) *HBB1* and *HBB2* are not, after all, 1:1 orthologues to snake *HBB1* and *HBB2*, respectively. Instead, both lizard β-like globins are more similar to snake *HBB2*, and, together with tuatara *HBB1*, form a monophylic sister clade to snake *HBB1* (Figure 2). The second tuatara β-globin (*HBB2* or $\beta^D$) groups outside this clade, suggesting that it is evolving at a very slow rate (Gorr et al., 1998). Its relationship with squamate *HBB1* and *HBB2* still remains unclear. More sequences and rigorous analyses are required to resolve their relationships.

The nomenclatures of β-like globin genes have become very inconsistent. I also propose renaming some of the reptilian β-globin chains to more accurately reflect their phylogenetic relationship. For example, based on the close relationship of lizard *HBB1* and *HBB2* to snake *HBB2*, I would suggest that lizard globins be renamed *HBB2-T1* and *HBB2-T2*. Likewise tuatara *HBB1* should be renamed *HBB2* and tuatara *HBB2* to *HBB1* once confirmation from more sequences of other species is available.

The presence of two β-like globins in tuatara and squamates indicates that these two globins existed in their common ancestor, suggesting a lepidosaur-specific duplication of an ancestral β-globin gene. In addition, my phylogeny suggests that the two β globins present in lizards are a result of lizard-specific duplication of snake-like *HBB2* gene, rather than squamate-specific duplication as had been originally proposed (Gorr et al., 1998).

In crocodiles and turtles, however, only one type of β-like globin gene has been discovered to date. This suggests that the reptile ancestor contained a single ancestral β-globin gene.

The absence of embryonic β-like globins in lizards reflects a different gene content and structure than the bird β-globin cluster of four genes. Chicken *HBE* and *HBR* are embryonic genes that are expressed in the early stages of development (Schalekamp and Van Goor, 1984, Baumann et al., 1987). Since all other jawed vertebrates that have been studied at the protein level possess both embryonic and adult expressed α- and β-like globin genes (Banville and Williams, 1985a, Goodman et al., 1987, Hoffmann and Storz, 2007), it would be surprising if lizards did not also possess at least one embryonic β-like globin gene. However, as for the embryonic α-like globin genes, there is no evidence for the existence of any embryonic β-like globin gene in the anole genome. Unless the absence of the embryonic *HBP* gene from lizards proves to be an assembly artefact, either *HBB1* or *HBB2* could be embryonically expressed in the anole and other lizards.

Unlike α-globins, the duplication of an ancestral β-globin gene into progenitors of early-expressed (*HBE*) and late-expressed (*HBB*) β-like globin genes is not clear for reptiles because of incomplete studies in different extant reptiles. Therefore, it still remains uncertain whether embryonic β-like globin gene existed in the common ancestor of birds and lizards, followed by independent divergence or lost in each of reptilian lineages. Alternatively, the duplication leading to early- and late-expressed β-like globin genes occurred independently in each of reptilian lineages. To solve these hypotheses, characterisation of the entire β-globin cluster and expression data are needed for various reptilian species.

**Evolution of β-globin genes in reptiles**

The inconsistency of β-globin gene copy numbers and conservation in different reptilian species makes it difficult to reconstruct the ancestral β-globin cluster at the stem of reptilian/mammalian divergence. Phylogenetic analysis of all reptilian β-globin amino acid sequences presented here showed species-specific clustering, suggesting that gene duplication occurred independently in each reptilian lineage.

Here, I propose two alternative models for the evolution of β-globin genes in reptiles, assuming that all β-globin genes have been identified. Since crocodiles and turtles contain one β-globin gene, the first model (Figure 4A) suggests that the reptilian ancestor contained a single ancestral β-globin gene, which first duplicated in the lepidosaur lineage (tuatara, snakes and lizards) to result in proto-$\beta^1$ and proto-$\beta^2$, both of whose descendants are seen in tuatara and snakes. However, during the evolution of lizards the proto-$\beta^1$ gene was secondarily lost from the genome, while proto-$\beta^2$ gene underwent tandem duplication to result in two paralogous β-globin genes in all lizards, as supported by the phylogenetic tree shown in Figure 2. In the avian lineage, there were further duplications of the ancestral β (independent from the duplication in lepidosaurs) to result in a four-gene cluster in chicken and zebrafinch. The exact number and timing of these duplications in the avian lineage is unknown at this stage because of incomplete β-globin studies in a range of birds.

Although no expression studies have been performed in lepidosaurs, the division of labour between early and adult stages of development could be similar to those of chickens and zebrafinch, but employ different gene copies. Thus, three independent duplications of β in the avian, lepidosaur and lizard lineages may have been followed by convergent evolution whereby selective pressure and purifying forces would have resulted in their genes having similar expression patterns and functions. As a consequence of independent duplications, the reptilian β-globin genes show fewer similarities to those of mammals.

The other possible, but less parsimonious, model (Figure 4B) proposes that the reptilian ancestor possessed two proto-β-globin genes, which underwent differential gene loss and duplication in different reptilian lineages. The proto-$\beta^1$ is found in crocodiles, turtles, tuatara, and snakes and duplicated further in birds, whereas in lizards it was secondarily lost from the genome. The proto-$\beta^2$ is found in tuatara and snakes, and as two paralogous copies in lizards, but was secondarily lost in the common ancestor of birds, crocodiles and turtles. According to this model, there may be either functional or remnants of proto-$\beta^2$ present in crocodiles and turtles and proto-$\beta^1$ in lizards.

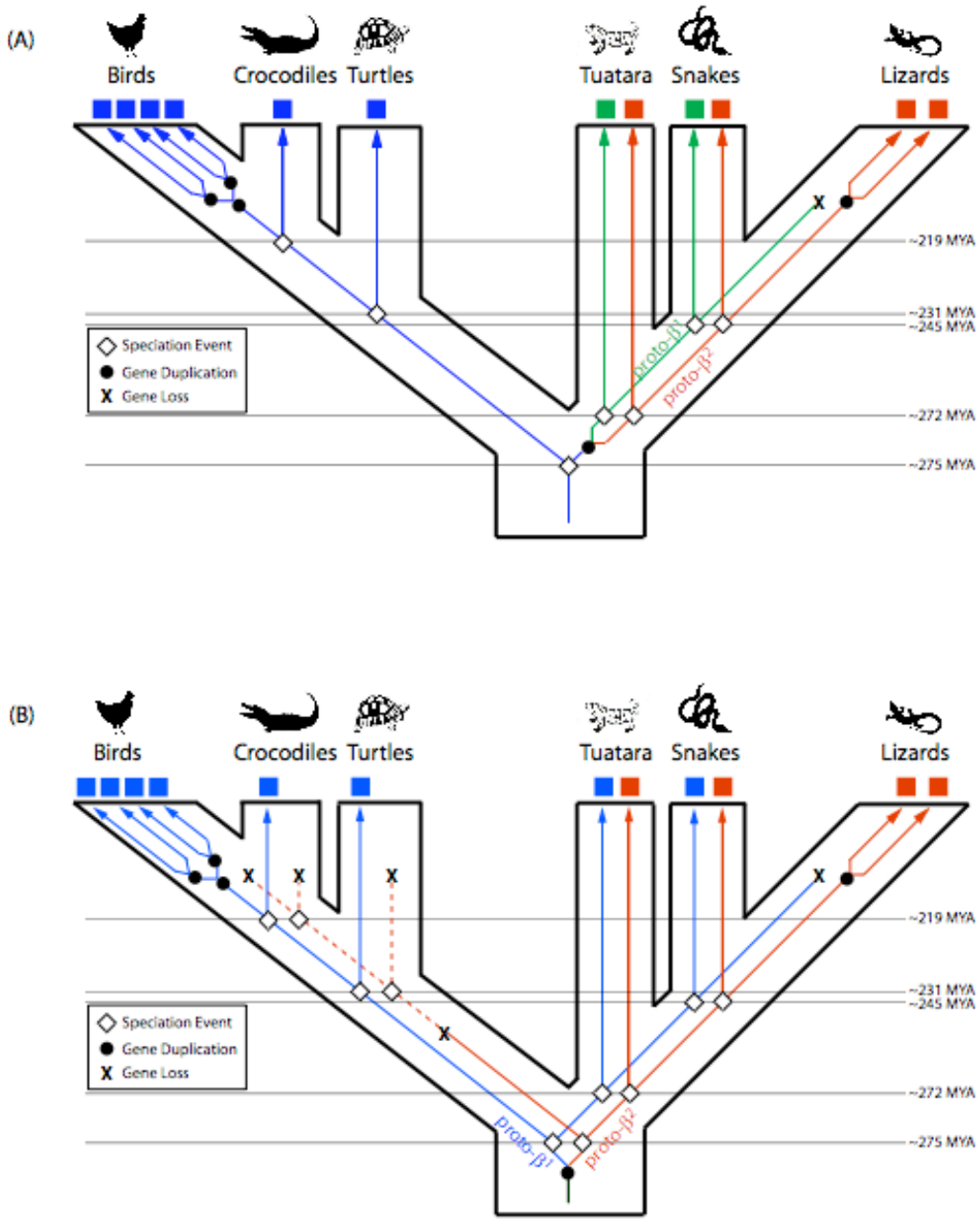



**Figure 4: Proposed models for the β-globin gene evolution in reptiles**

A) According to the *top* model, the reptilian ancestor contained a single ancestral β-globin gene that independently duplicated in the bird and lepidosaur (tuatara, snake and lizard) lineages. The latter duplication resulted in two proto-β genes; proto-$\beta^1$ (*green* branch) that is present in tuatara and snakes but lost in lizards, and proto-$\beta^2$ (*red* branch) that is present in tuatara and snakes and as two copies in lizards fulfil the species' requirement for oxygen; B) According to the *bottom* model, the reptilian ancestor already had two paralogous proto-β globin genes that underwent differential gene loss and diversification in different lineages. The proto-$\beta^1$ (*blue* branch) exists in all reptilian lineages (except lizards) and independently duplicates further in bird lineage. The proto-$\beta^2$ (*red* branch) however, is secondarily lost from the genome of the common ancestor of birds, crocodiles and turtles, but it is present in tuatara, snakes and further duplicated in lizards.

**Future study needed in another lizard species**

The anole genome browser shows the presence of *HBK* and *HBA* on two different scaffolds. Flanking analyses show that *HBK* is flanked by *MPG* and *C16orf35* at the 5' end and *GBY* at 3' end, which are typical in other tetrapods, suggesting a true α-globin cluster. The location of anole *HBA* (flanked by *ADCY9* and *GSG1L*) on another scaffold is unexpected and may be a product of incomplete and/or inaccurate assembly or another transposition event like β-globin.

The flanking genomic region of β-globin cluster also could not be established in the anole as the two β-like globin genes, and the genes that flank other amniote β-globin clusters (*RRM1*, *CCKBR* and *ILK*) lay on different scaffolds.

I therefore attempted to map BAC clones containing these genes onto anole chromosomes to determine the relationship of globin and their flanking genes. However, many repeated BAC library screening and FISH experiments failed to isolate BACs containing genes other the β-like globin genes, so it was not possible to map their locations. Even the mapping of two BAC clones containing the anole β-globin cluster was unsuccessful. Thus, it remains unknown if the anole α- and β-globin clusters and their flanking genes are located together or on two different chromosomes, as predicted by the transpositional model. As a consequence, the evolution of reptilian α- and β-globin clusters could not be determined from anole data.

The difficulties I described for mapping the anole α- and β-globin clusters were mostly due to poor quality BAC library and chromosome preparations being available for this species, a problem that could have been overcome if I used a species readily available in Australia. Thus, there is a need to study another lizard species to fill in the missing information and questions left unanswered by the anole. For this reason, I undertook studies in an Australian model lizard species, the bearded dragon lizard (*Pogona vitticeps*) for which resources such as a BAC library and chromosome preparations are readily available in Australia.

## Conclusion and future direction

The hypotheses presented here are still subject to much uncertainty as long as genome sequencing and assembly of the green anole is incomplete. It is vital to finish the analysis of this genome, and to add other reptile genomes at depth, to search for known embryonic globin genes. More DNA sequence data covering the entire α- and β- globin clusters from many different extant non-avian reptiles, along with their expression studies in different tissues and developmental stages, are needed to clarify α- and β-globin gene evolution and function in reptiles. The forthcoming sequencing projects of duck (*Anas platyrhynchos*) and turkey (*Meleagris gallopavo*) will be useful, but genome sequencing of a crocodile, turtle, tuatara and snake would provide much greater benefits to globin research. I therefore studied an Australian model reptile in order to delve further into the evolution of globin genes in reptiles.

## Acknowledgements

I thank Assoc. Prof. Travis Glenn (University of Georgia, USA) for providing the green anole gDNA, Dr Lee Miles (University of Sydney, Australia) for transporting them and Dr Dan Janes (Harvard University, USA) for providing preparations of green anole chromosomes. I also thank Hardip Patel (The Australian National University, Australia) for his help in building phylogenetic trees.

# PART B: GLOBIN GENE STRUCTURE IN A REPTILIAN SUPPORTS THE TRANSPOSITIONAL MODEL FOR AMNIOTE α- AND β-GLOBIN GENE EVOLUTION

***Authors:***

**Patel, V.S.**, Ezaz, T., Deakin, J.E., and Graves, J.A.M.

***Current status of paper (circle as appropriate):***

Planned/ In Preparation/ Submitted/ Under Revision/ Accepted/ Published

***Date paper accepted for publication or anticipated date of publication:***

21 October 2010

***Name of journal/book:***

Chromosome Research **18**: 897-907

***Extent to which research is your own:***

I designed and performed all experiments and analysed data in consultation with the co-authors.

***Your contribution to writing the paper:***

I wrote the draft of the manuscript, prepared figures and incorporated suggestions made by co-authors. I corresponded with the journal editor and responded to the referees' comments.

***Comments:***

Using the green anole lizard globin gene sequences (see section 4A) as reference for designing probes for the dragon lizard, I was able to isolate and map α- and β-globin clusters in the dragon lizard and study the evolution of globin genes in reptiles. I show that my observations support the transpositional model for the evolution of α- and β-globin genes in amniotes. The dragon lizard proved to be a better study model than the green anole lizard.

Chromosome Res (2010) 18:897–907
DOI 10.1007/s10577-010-9164-5

# Globin gene structure in a reptile supports the transpositional model for amniote α- and β-globin gene evolution

**Vidushi S. Patel · Tariq Ezaz · Janine E. Deakin · Jennifer A. Marshall Graves**

Received: 16 April 2010 / Revised: 20 October 2010 / Accepted: 21 October 2010 / Published online: 30 November 2010
© Springer Science+Business Media B.V. 2010

**Abstract** The haemoglobin protein, required for oxygen transportation in the body, is encoded by α- and β-globin genes that are arranged in clusters. The transpositional model for the evolution of distinct α-globin and β-globin clusters in amniotes is much simpler than the previously proposed whole genome duplication model. According to this model, all jawed vertebrates share one ancient region containing α- and β-globin genes and several flanking genes in the order *MPG-C16orf35-(α-β)-GBY-LUC7L* that has been conserved for more than 410 million years, whereas amniotes evolved a distinct β-globin cluster by insertion of a transposed β-globin gene from this ancient region into a cluster of olfactory receptors flanked by *CCKBR* and *RRM1*. It could not be determined whether this organisation is conserved in all amniotes because of the paucity of information from non-avian reptiles. To fill in this gap, we examined globin gene organisation in a squamate reptile, the Australian bearded dragon lizard, *Pogona vitticeps* (Agamidae). We report here that the α-globin cluster (*HBK, HBA*) is flanked by *C16orf35* and *GBY* and is located on a pair of microchromosomes, whereas the β-globin cluster is flanked by *RRM1* on the 3′ end and is located on the long arm of chromosome 3. However, the *CCKBR* gene that flanks the β-globin cluster on the 5′ end in other amniotes is located on the short arm of chromosome 5 in *P. vitticeps*, indicating that a chromosomal break between the β-globin cluster and *CCKBR* occurred at least in the agamid lineage. Our data from a reptile species provide further evidence to support the transpositional model for the evolution of β-globin gene cluster in amniotes.

**Keywords** Agamidae · evolution · haemoglobin · comparative mapping · lizard · *Pogona vitticeps*

**Abbreviations**

| | |
|---|---|
| α | Cluster of α-like globin genes |
| β | Cluster of β-like globin genes |
| BAC | Bacterial artificial chromosome |
| Blastn | Basic local alignment search tool nucleotide |
| Blastp | Basic local alignment search tool protein |
| BLAT | Blast-like alignment tool |
| bp | Base pairs |
| *C16orf35* | Chromosome 16 open read frame 35 |
| *CCKBR* | Cholecystokinin B receptor |
| dUTP | 2′-Deoxyuridine 5′-triphosphate |
| FISH | Fluorescence in situ hybridisation |

Responsible Editor: Wendy Bickmore.

V. S. Patel (✉) · J. E. Deakin · J. A. Marshall Graves
Research School of Biology,
The Australian National University,
Canberra, ACT 0200, Australia
e-mail: vidushi.patel@anu.edu.au

T. Ezaz
Institute for Applied Ecology, University of Canberra,
Bruce, ACT 2616, Australia



## Publication (Patel et al. 2010)

| | |
|---|---|
| *FTSJ1* | FtsJ homolog 1 (*E. coli*) |
| *GBY* | Globin Y |
| *HBA* | $\alpha^{A}$-Globin gene |
| *HBA-T3* | $\alpha^{A}$-Globin gene subunit 3 |
| *HBB* | β-Globin gene |
| *HBE* | ε-Globin gene |
| *HBK* | $\alpha^{D}$- Or μ- Globin gene |
| *HBP* | π-Globin gene |
| *HBQ* | θ-Globin gene |
| *HBW* | ω-Globin gene |
| *HBZ* | ζ-Globin gene |
| kb | Kilobase |
| *MPG* | *N*-methylpurine-DNA glycosylase |
| MY | Million years |
| MYA | Million years ago |
| NS | No sequences |
| nr/nt | Non-redundant nucleotide |
| *ORs* | Cluster of olfactory receptor genes |
| overgo | Overlapping oligonucleotides |
| *LUC7L* | *LUC7*-like |
| *RPS11* | Ribosomal protein S11 |
| *RRM1* | Ribonucleotide reductase M1 |

## Introduction

The progenitors of α- and β-globin genes evolved approximately 450 million years ago (MYA) from a single primordial globin gene (Goodman et al. 1975, 1987; Czelusniak et al. 1982). They further duplicated and diverged to form complex clusters with regulatory networks that coordinate the balanced expression of α-like and β-like globin genes to form different functional haemoglobin molecules at different stages of development (reviewed in Hardison 2001). Haemoglobins are haem-binding proteins that transport oxygen from outer respiratory surfaces to inner organs, to fulfil the demand for oxygen. In most teleost fishes (Chan et al. 1997; Gillemans et al. 2003; Pisano et al. 2003) and amphibians (Jeffreys et al. 1980; Hosbach et al. 1983; Fuchs et al. 2006), the α- and β-like globin genes are found together in one cluster, but in birds and mammals, they are grouped into distinct α-globin and β-globin clusters that are located on different chromosomes.

The evolutionary process that led to the separation of α- and β-globin genes in amniotes (mammals and reptiles, including birds) has been much debated. The original theory that a fission event separated duplicate genes (Jeffreys et al. 1980) was called into question by the discovery of a functional and seemingly ancient β-like globin gene (termed ω-globin, or *HBW*) in marsupials at the 3′ end of the α-globin cluster (Holland and Gooley 1997; Holland et al. 1998; De Leo et al. 2005; Cooper et al. 2006). Since this gene phylogenetically grouped as a sister clade to the avian β-globin genes, it was suggested that the amniote β-globin clusters arose via duplication (regional or whole genome) of the ancestral β-globin gene and diverged into the two major evolutionary lineages: one leading to a cluster of mammalian β-like globin genes and the other leading to the avian β-globin genes and the *HBW* (Wheeler et al. 2001, 2004). According to this model, the avian β-like globin genes are orthologous to the marsupial *HBW* gene, but both are paralogous to the mammalian β-like globin genes.

Since then, the model suggested by Wheeler et al. (2001, 2004) had been questioned, and there is now data from several sources to refute that *HBW* is orthologous to avian β-like globin genes (Aguileta et al. 2006; Opazo et al. 2008; Patel et al. 2008). Instead, Aguileta et al. (2006) and Opazo et al. (2008) hypothesised a new model for the evolution of β-globin clusters in amniotes where a duplication of an ancestral β-globin gene before the divergence of birds and mammals but after amniote/amphibian split resulted in two lineages, one copy of the gene leading to *HBW* and the other copy leading to all amniote β-globin clusters. In addition, results from Patel et al. (2008) showed strong support for a simple transpositional model (Jeffreys et al. 1980; Gillemans et al. 2003; Hardison 2005) for the evolution of α- and β-globin clusters in amniotes that seemed to be more parsimonious than all other models being proposed. It was observed that the genomic context of amniote α-globin clusters was similar to those of teleost and amphibian α–β globin clusters, but the β-globin clusters were completely different, being embedded in a group of olfactory gene receptors (Bulger et al. 1999; Flint et al. 2001; Burmester et al. 2002; Gillemans et al. 2003; Hardison 2005; Hughes et al. 2005; Patel et al. 2008).

According to the transpositional model, all amniote α-globin clusters derived from an ancient (ancestral) region containing *MPG-C16orf35-(α-β)-GBY-LUC7L* that has been conserved since the radiation of jawed

vertebrates (410 MYA), except for loss in some species of genes *HBW* and/or *GBY* (a member of the globin family; Flint et al. 2001; Hughes et al. 2005; Patel et al. 2008). However, the amniote β-globin cluster evolved in an ancestral amniote just before divergence of reptiles and mammals (310 MYA) by transposition of a single copy of the β-globin gene from this ancient region and insertion into another region containing many olfactory receptor genes (*ORs*), flanked by cholecystokinin B receptor (*CCKBR*) at the 5′ end and ribonucleotide reductase M1 (*RRM1*) at the 3′ end (Bulger et al. 1999; Gillemans et al. 2003; Hardison 2005, 2008; Patel et al. 2008). It then duplicated and diverged further in all lineages to form a dominant and functional β-globin cluster. According to this model, all amniotes should therefore possess two distinct globin clusters on different chromosomes. One cluster is *MPG-C16orf35-(α-β)-GBY-LUC7L,* comprising many active α-like globin genes and retaining β (*HBW*) and/or *GBY* in most species. The other is *CCKBR-ORs-(β)-ORs-RRM1*, comprising many active β-like globin genes embedded in an olfactory gene cluster.

The transpositional model, however, lacks support at the stem of amniote radiation more than 300 MYA when reptiles diverged from the line leading to mammals. There is little information on gene organisation, genomic context and location of α- and β-globin in non-avian reptiles, since research to date has focussed on birds and mammals. Amino acid sequences of haemoglobins in some non-avian reptiles are available from early studies (Gorr et al. 1998 and references therein). Although the genome of a green anole lizard, *Anolis carolinensis*, has been sequenced, assembly around the α- and β-globin clusters is currently incomplete. Hoffmann et al. (2010b) found two α-like globin genes and two β-like globin genes, all located on different scaffolds (*HBK* on scaffold 2790, *HBA* on 1188, *HBB1* on 7008, and *HBB2* on 3777). The fragmentation of the current assembly did not permit discovery of other putative globin genes (such as *HBW* and embryonic *HBP* and *HBE*) in these clusters, nor establishment of the arrangement of these genes, so it could not be determined whether reptiles possess distinct α-globin and β-globin clusters like birds and mammals, or only one α-β globin cluster like amphibians and teleosts.

Identification of the α-globin and β-globin clusters, their flanking genes and establishing their relative locations in a non-avian reptile is therefore critical for testing the transpositional model for the evolution of amniote α- and β-globin clusters. We selected an Australian dragon lizard, *Pogona vitticeps* (squamate, Agamidae) for a comparative study on globin genes. Although its genome has not yet been sequenced, this species has been selected as a model reptilian species for comparative genomics studies (Ezaz et al. 2009), and genomic DNA, tissues, metaphase chromosome preparations and a bacterial artificial chromosome (BAC) library are available.

The aims of our study were to establish the genomic organisation of the α- and β-globin clusters and their flanking genes in the dragon lizard, *P. vitticeps*, and map their locations in the genome, in order to test the transpositional model for the evolution of globin genes. We therefore screened the *P. vitticeps* BAC library for α-globin and β-globin genes and genes that flank each cluster in mammals and birds. We used the green anole whole genome sequences to design probes for identification, isolation and physical mapping of these genes in the dragon lizard. Our study revealed that the dragon lizard, *P. vitticeps*, possesses two distinct clusters of α-globin and β-globin genes that are located on different chromosomes. The composition of the dragon lizard's globin gene clusters is consistent with the transpositional theory of evolution of amniote globin genes.

## Materials and methods

### Overgo design for globin and flanking genes

Since sequence information for globin and their flanking genes were not available for *P. vitticeps*, overgos for BAC library screening were designed using the green anole whole genome sequences (anoCar1: Feb 2007; http://genome.ucsc.edu). Eight genes (*C16orf35, HBK, HBA, GBY, CCKBR, HBB1, HBB2* and *RRM1*) were found on different scaffolds (Table 1), and their sequences were predicted using gene prediction programmes such as Genscan (Burge and Karlin 1997) and Genomescan (Yeh et al. 2001) using default settings. Upon confirmation of the predicted gene sequences by Blastn and Blastp searches against the nr database (http://blast.ncbi.nlm.nih.gov/Blast.cgi/), overgos for each gene were

**Table 1** List of overgos designed from the predicted anole globin and flanking genes found on different scaffolds to screen *P. vitticeps* BAC library

| Genes | Green anole scaffold ID | Forward overgo | Reverse overgo |
|---|---|---|---|
| *HBA* | 1188 | GACCTCCATGCCGAGAAGCTCCGT | AGTTGACAGGGTCAACACGGAGCT |
| *HBK* | 2790 | GCTGATCAAGTGCTTCCACGTGGT | CAGGTGAGTGGCCAGAACCACGTG |
| *GBY-1* | 2790 | GCGGCACATCCGGGAAATCTGGAC | ATTTTCAAAGGCCGCGGTCCAGAT |
| *GBY-2* | 2790 | TTCAGCGACTACCCTGCCAGCAAG | CCGTCTTGAAGTACTGCTTGCTGG |
| *C16orf35* | 2790 | AGTGATGACATGACGCTCACCAGC | AGTTGTCCATGCTGGGGCTGGTGA |
| *HBB-1/2* | 7008 and 3777 | GGACAACATCAAGGACACCTTCGC | CAGCTCGCTCAGCTTGGCGAAGGT |
| *RRM1* | 985 | AAGATCACCTCCCGGATCCAGAAG | TAAGACCGTAGCACAGCTTCTGGA |
| *CCKBR* | 1207 | TCACCAACTCCTTCCTGCTCTCGC | AGGTCGCTGAGGGCCAGCGAGAGC |

designed using OvergoMaker (Washington University Genome Sequencing Centre 2010). For *HBB1* and *HBB2*, rather than designing two pairs of overgos for two genes, only one pair of overgos was designed from the conserved region between them. For *GBY*, two pairs of overgos were designed, one from its exon 2 and another from exon 3, to increase the chances of hybridisation. Therefore, a total of eight overgo pairs based on the green anole sequences were designed for cross-species screening using *P. vitticeps* BAC library (Table 1).

### *P. vitticeps* BAC library screening

High-density filters (~6.2× coverage) from a *P. vitticeps* genomic BAC library (commissioned via Amplicon Express; http://www.genomex.com/) were screened with eight radioactive labelled overgo probes according to the protocol as described in Ezaz et al. (2009). The presence of globin and/or flanking genes in the BAC clones was confirmed by dot blots, following protocol described in Patel et al. (2008). In addition, BAC clones were fingerprinted by the British Columbia Cancer Agency (Vancouver, Canada) to identify sizes and overlapping clones.

### Confirmation of gene content of BAC clones by direct sequencing

To further confirm that the isolated *P. vitticeps* BAC clones contain homologues of globin genes and their flanking genes, sequencing was performed using overgos (Table 1) as sequencing primers following the protocol described in Deakin et al. (2008). Some extra primers (designed from anole sequences) were also used to increase the success of direct sequencing. They were: *HBA* 'TGCTGACCGCTGAGGACCG CAAAC', *HBK* 'AAGTTGACTGGGTCCACGCG GAGG', *C16orf35* 'AATTCTGGCCACCAAGTCG GACAT', *HBB* 'AAGGCGTGGGACACCACGTT GACC', *RRM1* 'ATTGACCAGAGCCAGTCCCT GAAC' and 'GCTCAGCAATGTGGATGTT CAGGG'. Sequences obtained were then used for Blastn searches against the nucleotide collection (nr/nt) database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and also for BLAT searches against the green anole genome sequence database (anoCar1: Feb 2007; http://genome.ucsc.edu/cgi-bin/hgBlat). The percent identities revealed by Blastn and BLAT searches were calculated over the short region of alignment and not the whole sequence. These nucleotide sequences have been deposited in the NCBI Genbank database (accession numbers HM038110-HM038117).

### Cell culture, chromosome preparations and FISH mapping of globin genes

Cell cultures and chromosome preparations of *P. vitticeps* were performed following the protocol described in Ezaz et al. (2005). The verified BAC clones containing globin and/or flanking genes were labelled with fluorchrome-labelled dUTPs (spectrum green and orange) and hybridised to *P. vitticeps* chromosomes using two-colour fluorescence *in situ* hybridisation (FISH), following protocols described in Deakin et al. (2008) and Ezaz et al. (2009).

## Results

### Isolation and identification of *P. vitticeps* BAC clones

Nine *P. vitticeps* BAC clones containing homologues of globin genes and/or flanking genes (*C16orf35, HBA, HBK, GBY, CCKBR, HBB, RRM1*) were isolated using eight overgo pairs and were verified by dot blot (Table 2).

Three *P. vitticeps* BAC clones (Pv29M9, Pv176F6 and Pv236C5) were found to contain at least one α-like globin gene (*HBK*) and a flanking gene *C16orf35* (Table 2). Two of these BAC clones (Pv176F6, Pv236C5) contained an additional α-like globin gene (*HBA*). We also found an orthologue of *GBY* gene in one of the clones (Pv236C5) that contained two α-like globin genes (*HBK* and *HBA*), as well as their flanking genes *C16orf35* and *GBY*. Based on the α-globin gene arrangement in other jawed vertebrates, we predicted that the four genes are arranged in the order as *C16orf35-HBK-HBA-GBY*. BAC fingerprinting results showed that all three BAC clones (Pv29M9, Pv176F6 and Pv236C5) overlapped with each other, covering more than 265Kb of α-globin and its flanking region (Fig. 1).

We performed direct sequencing of BAC clones and subsequent sequence analysis using Blastn searches against the whole nr/nt database as well as BLAT searches against the green anole genome sequence database to verify the presence of these genes. The sequences contained a portion of the exon to which the overgo/sequencing primer was designed and the adjacent intron. Since all α- and β-globin genes are encoded by three short exons (for example, the β-like globin gene has exons 92, 223 and 129 bp), the alignment matches to these sequences were also short and restricted to the exonic sequence (data not shown). Sequencing confirmed the presence of *C16orf35, HBK* and *HBA* in Pv176F6, providing further evidence that all overlapping BAC clones contain the α-globin cluster and flanking genes (Table 2).

Our search for genes in the β-globin cluster yielded only one BAC clone (Pv124N8) containing a *HBB* gene. We could not confirm if this gene was a homologue of the green anole *HBB1* or *HBB2,* as the overgo was designed from the region conserved between these two genes. Even the sequence data could not distinguish the two genes because *P. vitticeps* sequences aligned to both green anole *HBB* sequences with similar identity scores. For example, *P. vitticeps* partial *HBB* fragment 2 sequence (Genbank HM038111) obtained 84% identity score over the aligned region of 107 bp to green anole *HBB2* and 82.6% identity score over the aligned region of 110 bp to green anole *HBB1*.

We searched for *CCKBR* and *RRM1*, the genes that flank the β-globin clusters of other amniotes, using dot blots and BAC fingerprinting (Fig. 1) in *P. vitticeps*. The *HBB*-containing BAC clone (Pv124N8) contained neither of these flanking genes. Instead, three separate

**Table 2** Genes present in the dragon BAC clones and their respective chromosomal locations and sizes

| Clone ID | Gene(s) identified | Chromosomal location | Size (bp) | Genbank accession numbers[a] |
|---|---|---|---|---|
| Pv 29M9 | *C16orf35, HBK* | Microchromosome | 156,945 | NS[b] |
| Pv 176F6 | *C16orf35, HBK, HBA* | Microchromosome | 168,337 | HM038117 |
| | | | | HM038114 |
| | | | | HM038115 |
| Pv 236C5 | *C16orf35, HBK, HBA, GBY* | Microchromosome | 181,566 | NS[b] |
| Pv 124N8 | *HBB* | 3q | 184,002 | HM038110-HM038113 |
| Pv 61D8 | *RRM1* | 3q | 173,277 | HM038116 |
| Pv 140I9 | *RRM1* | 3q | 101,816 | HM038109 |
| Pv 131D7 | *RRM1* | 3q | 82,293 | HM038116 |
| Pv 179I7 | *CCKBR* | 5p | 223,285 | NS[b] |
| Pv 239F4 | *CCKBR* | 5p | 121,150 | NS[b] |

BAC size was determined from fingerprint results

[a]Genbank accession numbers for genes present in the BAC clones

[b]No sequences were retrieved from these BAC clones

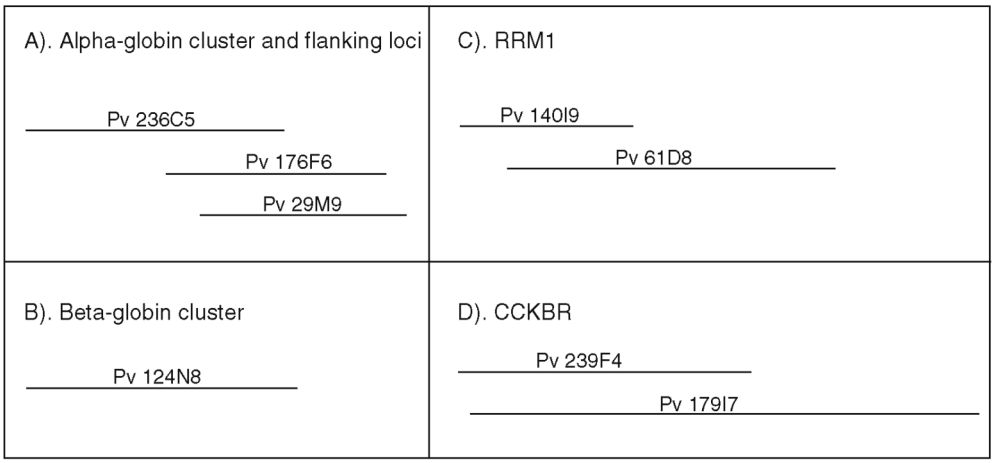



**Fig. 1** Contig map generated by BAC fingerprinting. Eight out of nine clones sent for BAC fingerprinting contigged into four major scaffolds: **a** α-globin cluster and the flanking loci; **b** β-globin cluster; **c** *RRM1*; and **d** *CCKBR*

clones (Pv61D8, Pv140I9 and Pv131D7) were identified and confirmed to contain *RRM1* and two separate clones (Pv179I7 and Pv239F4) were identified to contain *CCKBR* (Table 2). BAC fingerprinting showed that two of the three *RRM1*-containng BAC clones (Pv61D8 and Pv140I9) overlap, extending the region up to 198 Kb (Fig. 1). The third clone (Pv131D7) was also confirmed to contain partial *RRM1* sequence, which was identical to the sequence retrieved from Pv61D8 (Genbank HM038116) and mapped to the same chromosomal location as Pv61D8 and Pv140I9 (Table 2). However, no overlapping was observed between Pv131D7 and the other two *RRM1*-containing BAC clones by the BAC fingerprinting software, perhaps because of its small size. Furthermore, the two overlapping BAC clones (Pv179I7 and Pv239F4) identified as containing *CCKBR* by dot blots span more than 200 kb. However, we could not confirm that they contain *CCKBR* by sequencing, probably because of non-specificity of the primers (which were designed from the anole) binding to the region in *P. vitticeps*. None of these five BAC clones overlaps with Pv124N8, suggesting that there is perhaps a small gap between the BAC clones for these genes.

### Chromosomal location of α- and β-globin clusters and flanking genes in *P. vitticeps*

To identify the locations of the BAC clones containing the two globin gene clusters and their flanking genes in the dragon lizard, two-colour FISH was performed (Fig. 2).

We found that the α-globin and β-globin clusters were located on different chromosomes. All three BAC clones containing the α-globin cluster co-localised on a pair of microchromosomes (Fig. 2a), whereas the BAC clone containing the β-globin cluster is located on the long arm of chromosome 3 (3q) in dragon lizard (Fig. 2b).

We also performed two-colour FISH using three BAC clones containing *HBB*, *CCKBR* and *RRM1* to determine whether these genes lie in close proximity. We found that the *RRM1* clone resides on 3q close to the β-globin cluster. Two-colour FISH produced two signals that were overlapping (Fig. 2c), indicating that *RRM1* is very close to the *HBB* gene in *P. vitticeps*, as would be expected for a flanking gene.

However, we found that *CCKBR* was not located on chromosome 3 with *HBB* and *RRM1*, but is located on the short arm of chromosome 5 (5p) (Fig. 2d), therefore, *CCKBR* gene does not flank the β-globin cluster in the dragon lizard.

## Discussion

Globin gene evolution is a classic example of gene duplication and transposition leading to gene shuffling and evolutionary innovation. Many different and increasingly complex models were proposed to

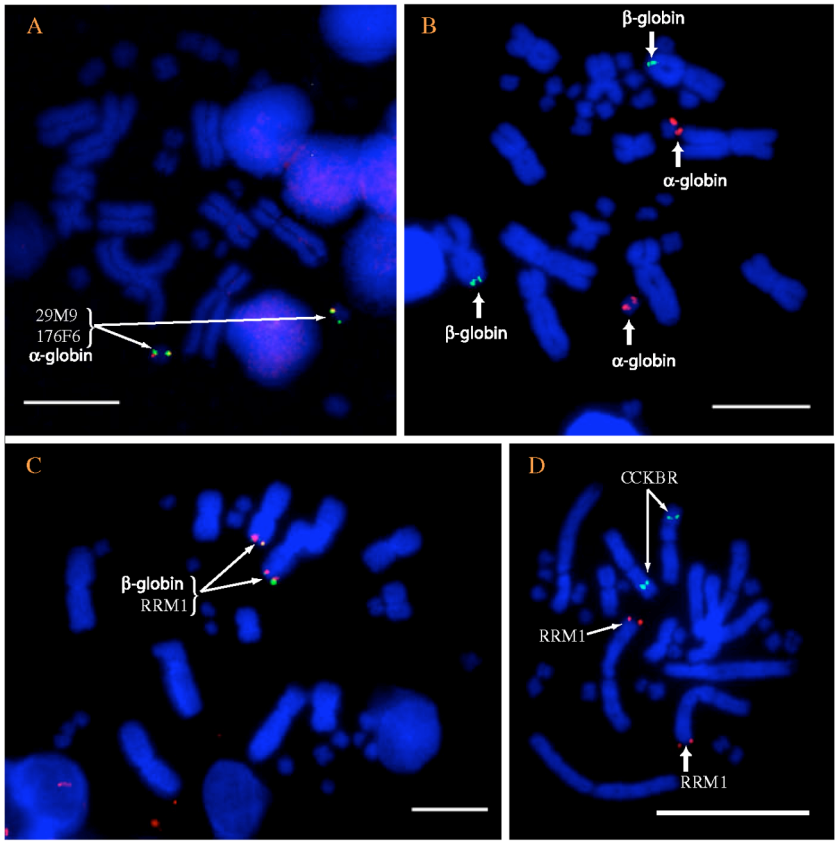



**Fig. 2** Localisation of α- and β-globin clusters and their flanking genes in the Australian dragon lizard, *P. vitticeps*. **a** BAC clones Pv176F6 (*green*) and Pv235C5 (*red*) containing α-globin cluster co-localised on a pair of microchromosomes. The colour of signals appears *yellow* where probes overlap; **b** BAC clone Pv236C5 containing α-globin cluster (*HBK*, *HBA*) and flanking genes *C16orf35* and *GBY* (*red*) is located on a pair of microchromosomes whereas BAC clone Pv124N8 containing β-globin cluster (*green*) is on 3q; **c** BAC clone Pv124N8 containing the β-globin cluster (*red*) is located on 3q, as is BAC clone Pv131D7 containing *RRM1* (*green*); and **d** BAC clone Pv140I9 containing *RRM1* (*red*) is located on 3q whereas BAC clone Pv239F4 containing *CCKBR* (*green*) is on 5p bars equal 10μM

explain the evolution of α- and β-globin gene clusters in vertebrates (Jeffreys et al. 1980; Wheeler et al. 2001, 2004; Aguileta et al. 2006; Opazo et al. 2008). However, the transpositional model is the simplest and most parsimonious model proposed to explain globin gene evolution. In this model, the ancestral globin gene cluster contained both α- and β-globin genes (duplicated from a single ancestral globin gene) in the genomic arrangement *MPG-C16orf35-(α-β)-GBY-LUC7L*, which is retained in amniotes with the loss of most or all the β-like globin genes (Flint et al. 2001; Hughes et al. 2005; Patel et al. 2008). In amniotes a separate β-globin gene cluster arose via transposition of a copy of the β-globin gene into a cluster of *ORs*, generating the region containing *CCKBR-ORs-(β)-ORs-RRM1* that is observed in all sequenced amniote genomes (Bulger et al. 1999; Gillemans et al. 2003; Hardison 2005, 2008; Patel et al. 2008).

At the time this model was proposed, no genome sequences were available for any non-avian reptiles. Since then, the genome of the green anole lizard was sequenced, but the vital information required to further test the transpositional model was missing from its genome assembly. We therefore isolated and mapped globin clusters and their flanking genes in an Australian model reptile, the dragon lizard *P. vitticeps*, to determine whether there were two separate clusters of

α- and β-globin genes in the same genomic context as in other amniotes, as predicted by the transpositional model. While this paper was in preparation, Hoffmann et al. (2010b) published an analysis of globin genes from several sequenced vertebrates, including the green anole, finding lineage-specific differences in some of the globin genes; however, the anole assembly is, as yet, too incomplete to determine globin gene arrangement in this species.

Comparisons of our data with analyses of sequence from the green anole and birds enabled us to deduce the arrangement of the α- and β-globin gene clusters in an ancestral amniote, providing further evidence to support previously reported transpositional model of globing gene evolution in amniotes.

## Evolution of the α-globin cluster in amniotes

Our study revealed that two α-globin genes *HBK* and *HBA* were contained in a single BAC clone, which was located on a microchromosome. This implies that at least two α-like globin genes are present in the α-globin cluster of *P. vitticeps*. However, in green anole only *HBK* was present between the flanking markers *C16orf35* and *GBY*, whereas *HBA* was located on a different scaffold (Hoffmann et al. 2010b). This could be an artefact due to the incomplete assembly of this region in anole (Hoffmann et al. 2010b), although it is possible that *HBA* has relocated onto another region, providing another example of lineage-specific rearrangement of the globin locus similar to that which we proposed for the origin of β-globin. Without mapping data, we cannot tell if these scaffolds are adjacent.

It is also plausible that more α-like globin genes, such as early expressed embryonic *HBP* (referred as *HBZ* in mammals), are present within the anole α-globin cluster, but were absent from the incomplete assembly from the anole genome. Since sequence information for this gene was not available from the anole or any other non-avian reptiles, we were unable to design probes for BAC library screening so we cannot determine whether *HBP* is present in *P. vitticeps*. To our knowledge, no embryonic globin genes have ever been described in non-avian reptiles, previous studies having focussed on isolating adult haemoglobin molecules from the blood of an adult non-avian reptile (Gorr et al. 1998). Thus, there is no direct evidence for embryonic α-globin genes in non-avian reptiles. However, since the progenitor of *HBP* was present in the common ancestor of tetrapods and it has been retained in all other descendant lineages (Hoffmann and Storz 2007), embryonic globins may well be discovered in the genomes of non-avian reptiles by more complete sequencing and assembly of the anole and dragon lizard globin clusters. Alternatively, *HBP* may have indeed been lost from some non-avian reptiles, and instead, the *HBK* gene was recruited to fulfil their embryonic requirement of oxygen. New expression studies are required to test this hypothesis (Hoffmann et al. 2010b).

We found that this cluster of α-globin genes in the dragon lizard is flanked by *C16orf35* on one side and *GBY* on the other. Hoffmann et al. (2010b) observed that *HBK* is flanked by the same genes in the green anole lizard, which diverged from agamid lizards ~144.2 MYA (Hedges and Vidal 2009). This suggests that all lizards possess the ancestral *C16orf35-(α)-GBY* arrangement. We were unable to isolate clones containing the more distant flanking markers *MPG* and *LUC7L,* but *MPG* is present 5′ to the green anole lizard *C16orf35*, so probably flanks the cluster in all lizards. *LUC7L* is located on a different anole scaffold (839: http://genome.ucsc.edu/). It would be advantageous to map the two scaffolds to determine whether *LUC7L* flanks the α-globin cluster as it does in other birds, mammals (except *Mus musculus*) and amphibians. Our data are therefore consistent with the hypothesis that the ancestor of all reptiles and birds contained the α-globin cluster in the genomic context *5′-MPG-C16orf35-α(HBP-HBK-HBA)-GBY-LUC7L-3′*.

There remains a transcriptionally active β-like *HBW* gene within the α-globin cluster of non-eutherian mammals, located immediately after *HBA-T3* (in monotremes) or *HBQ* (in marsupials; Wheeler et al. 2001, 2004; De Leo et al. 2005; Cooper et al. 2006; Patel et al. 2008). No *HBW* is present in eutherian genomes (Hoffmann et al. 2008; Opazo et al. 2008, 2009). We could not find any trace of an *HBW* orthologue in the dragon lizard, nor in the green anole, zebrafinch (*Taenippygia guttata*) and chicken (*Gallus gallus*) sequence databases. However, the presence of β-globin genes adjacent to α-globin in amphibians and teleost fishes and the phylogenetic position of *HBW* as an outgroup to bird and mammal β-globin genes suggests that it diverged from amniote β-like globin genes before the divergence of amniotes (Aguileta et al. 2006; Opazo et al. 2008; Patel et al. 2008), but was pseudogenised or lost in the ancestor of reptiles. This makes less parsimonious the alterna-

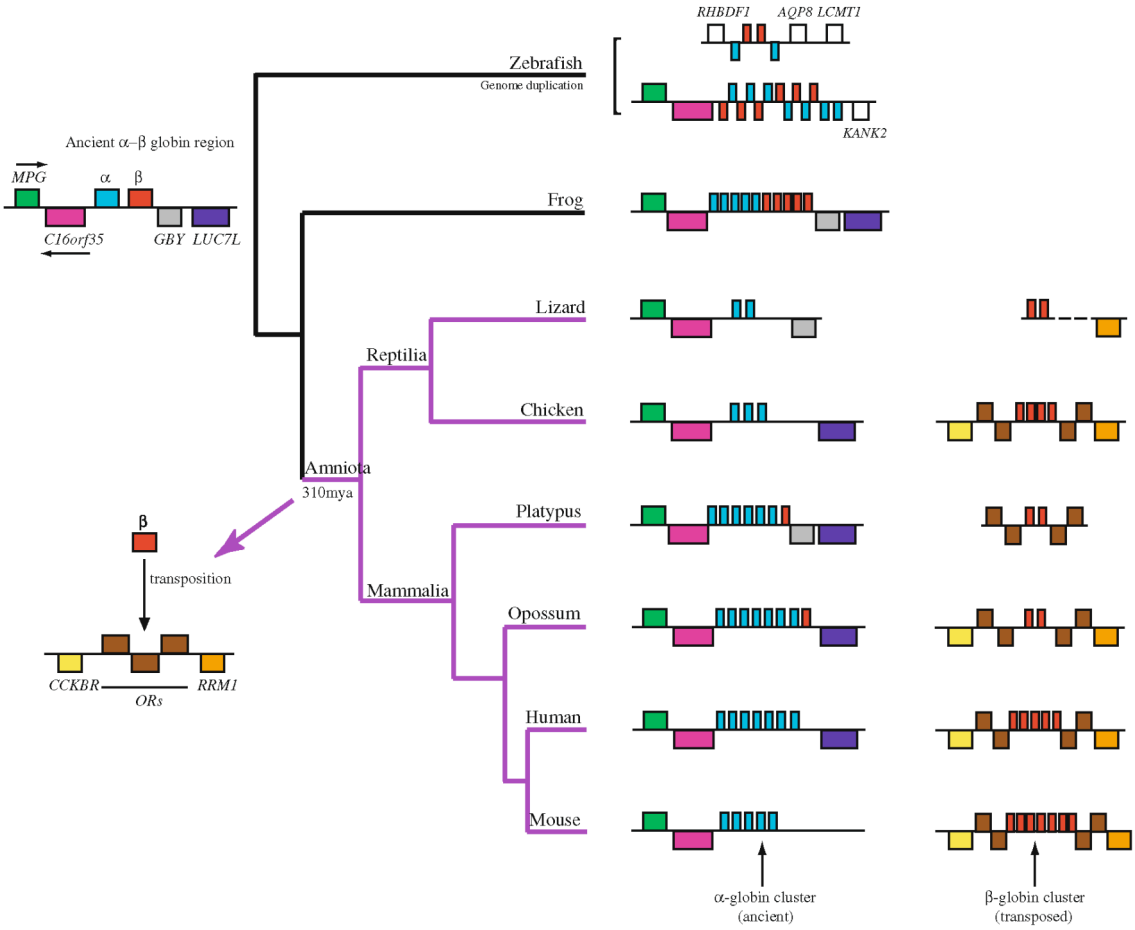



**Fig. 3** Transposition model for the evolution of α- and β-globin clusters in amniotes. Position of genes (represented by coloured *boxes* as described in the *left panels*, not to scale) in and around globin clusters in different vertebrates. Presence of the same markers (*MPG, C16orf35* and *LUC7L*) flanking the α-β-globin region in frogs and fish (one duplicate), and the α-globin cluster of amniotes (including dragon lizard), implies orthology. In mouse only, *LUC7L* has been separated from the α-globin cluster. The presence of a separate β-globin cluster flanked by olfactory receptor genes, *CCKBR* and *RRM1* in all amniotes implies transposition to a new site before the amniote divergence

tive hypothesis that *HBW* evolved in the mammalian ancestor after their divergence from reptiles, but was lost again in eutherian mammals. Investigation of the globin clusters in other reptiles (such as in turtle, crocodile or tuatara) could test this interpretation.

The globin-related gene *GBY* lies beside the α-globin cluster of monotremes (Patel et al. 2008) as well as the ancient α-β globin cluster of amphibians (Fuchs et al. 2006). It is an ancient duplicate of the globin gene that seems to be present in the common ancestor of all vertebrates and was independently lost in different lineages (Hoffmann et al. 2010a). Our finding of *GBY* within the α-globin cluster of the dragon lizard confirms that this gene was present in an ancestral amniote and was lost secondarily from therian mammal and bird genomes. Its precise function is worth investigating in the species possessing *GBY*.

Comparative data therefore support the hypothesis that the *5′-MPG-C16orf35-α(HBP-HBK-HBA)-β(HBW)-GBY-LUC7L-3′* arrangement was present before the divergence of all contemporary reptiles and mammals (310 MYA), but *HBW* was lost independently in birds, squamates (or their common

ancestor) and eutherian mammals, and *GBY* in therian mammals and birds.

### Evolution of the β-globin cluster in amniotes

Since in chicken and zebrafinch, as well as mammals, β-globin clusters are embedded in olfactory genes and flanked by *RRM1* and *CCKBR* (Bulger et al. 1999), we deduced that this arrangement was ancestral and predicted that it would be present also in non-avian reptiles.

It was difficult to find the gene content and flanking genes of the dragon lizard β-globin cluster. One *HBB* gene was located on the long arm of chromosome 3 and mapped adjacent to a BAC containing the flanking gene *RRM1*. However, *CCKBR* does not flank *HBB* on the other side, but is located on another chromosome (5p), suggesting a fission event that resulted in re-location to another chromosome. There are few useful data from the current assembly of the green anole, since two adult β-like globin genes are located on two different scaffolds (Hoffmann et al. 2010b) and genes *CCKBR* and *RRM1* also on separate scaffolds. The non-overlapping BACs in the dragon lizard and the incomplete assembly of the anole genome sequence leave the possibility that lizards possess more β-like globin genes.

Since β-globin clusters are flanked by *RRM1* and *CCKBR* in birds as well as mammals, it is highly likely that 5′-*CCKBR-ORs-(β)-ORs-RRM1*-3′ was present in the common ancestor of all amniotes and was disrupted either in the dragon lizard lineage or a lizard ancestor.

## Conclusion

Our study of globin genes in the dragon lizard, *P. vitticeps*, has brought us a step closer in understanding the gene structure, arrangement and evolution of α- and β-globin clusters in amniotes. Our observation that, like birds and mammals, the dragon lizard, *P. vitticeps*, possesses two distinct globin clusters located within different genomic contexts provides strong support for the transpositional model for the evolution of globin clusters in amniotes (Fig. 3; Patel et al. 2008). However, further information is needed from other non-avian reptiles to gain a wider view of the arrangement and evolution of reptilian α- and β-globin clusters. A well-assembled non-avian reptile genome sequence is required (e.g., turtle, crocodile and tuatara) if we are to delve deeper into the evolution of these important and iconic genes.

**Acknowledgements** This study was partially supported by an Australian Research Council Discovery grant (ARC DP0881196) awarded to Stephen Sarre, Arthur Georges and Scott Edwards, and a block grant from the Research School of Biology at the Australian National University (ANU). VSP is supported by ANU PhD Gradate School Scholarship.

## References

Aguileta G, Bielawski JP, Yang Z (2006) Evolutionary rate variation among vertebrate beta globin genes: implications for dating gene family duplication events. Gene 380:21–29

Bulger M, van Doorninck JH, Saitoh N et al (1999) Conservation of sequence and structure flanking the mouse and human beta-globin loci: the beta-globin genes are embedded within an array of odorant receptor genes. Proc Natl Acad Sci USA 96:5129–5134

Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. J Mol Biol 268:78–94

Burmester T, Ebner B, Weich B et al (2002) Cytoglobin: a novel globin type ubiquitously expressed in vertebrate tissues. Mol Biol Evol 19:416–421

Chan FY, Robinson J, Brownlie A et al (1997) Characterization of adult alpha- and beta-globin genes in the zebrafish. Blood 89:688–700

Cooper SJ, Wheeler D, De Leo A et al (2006) The mammalian alphaD-globin gene lineage and a new model for the molecular evolution of alpha-globin gene clusters at the stem of the mammalian radiation. Mol Phylogenet Evol 38:439–448

Czelusniak J, Goodman M, Hewett-Emmett D et al (1982) Phylogenetic origins and adaptive evolution of avian and mammalian haemoglobin genes. Nature 298:297–300

De Leo AA, Wheeler D, Lefevre C et al (2005) Sequencing and mapping hemoglobin gene clusters in the Australian model dasyurid marsupial Sminthopsis macroura. Cytogenet Genome Res 108:333–341

Deakin JE, Koina E, Waters PD et al (2008) Physical map of two tammar wallaby chromosomes: a strategy for mapping in non-model mammals. Chromosome Res 16:1159–1175

Ezaz T, Quinn AE, Miura I et al (2005) The dragon lizard *Pogona vitticeps* has ZZ/ZW micro-sex chromosomes. Chromosome Res 13:763–776

Ezaz T, Moritz B, Waters P et al (2009) The ZW sex microchromosomes of an Australian dragon lizard share no homology with those of other reptiles or birds. Chromosome Res 17:965–973

Flint J, Tufarelli C, Peden J et al (2001) Comparative genome analysis delimits a chromosomal domain and identifies key regulatory elements in the alpha globin cluster. Hum Mol Genet 10:371–382

Fuchs C, Burmester T, Hankeln T (2006) The amphibian globin gene repertoire as revealed by the *Xenopus* genome. Cytogenet Genome Res 112:296–306

Gillemans N, McMorrow T, Tewari R et al (2003) Functional and comparative analysis of globin loci in pufferfish and humans. Blood 101:2842–2849

Goodman M, Moore GW, Matsuda G (1975) Darwinian evolution in the genealogy of haemoglobin. Nature 253:603–608

Goodman M, Czelusniak J, Koop BF et al (1987) Globins: a case study in molecular phylogeny. Cold Spring Harb Symp Quant Biol 52:875–890

Gorr TA, Mable BK, Kleinschmidt T (1998) Phylogenetic analysis of reptilian hemoglobins: trees, rates, and divergences. J Mol Evol 47:471–485

Hardison RC (2001) New views of evolution and regulation of vertebrate beta-like globin gene clusters from an orphaned gene in marsupials. Proc Natl Acad Sci USA 98:1327–1329

Hardison RC (2005) Globin genes: evolution. Encylcopedia of life sciences. Wiley, Chichester. doi:10.1038/npg.els.0005134, http://www.els.net/

Hardison RC (2008) Globin genes on the move. J Biol 7:35

Hedges SB, Vidal N (2009) Lizards, snakes, and amphisbaenians (Squamata). In: Hedges SB, Kumar S (eds) The timetree of life. Oxford University Press, pp 383–389

Hoffmann FG, Storz JF (2007) The {alpha}D-globin gene originated via duplication of an embryonic {alpha}-like globin gene in the ancestor of tetrapod vertebrates. Mol Biol Evol 24:1982–1990

Hoffmann FG, Opazo JC, Storz JF (2008) Rapid rates of lineage-specific gene duplication and deletion in the alpha-globin gene family. Mol Biol Evol 25:591–602

Hoffmann FG, Opazo JC, Storz JF (2010a) Gene cooption and convergent evolution of oxygen transport hemoglobins in jawed and jawless vertebrates. Proc Natl Acad Sci USA 107:14274–14279

Hoffmann FG, Storz JF, Gorr TA et al (2010b) Lineage-specific patterns of functional diversification in the {alpha}- and {beta}-globin gene families of tetrapod vertebrates. Mol Biol Evol 27:1126–1138

Holland RA, Gooley AA 2nd (1997) Characterization of the embryonic globin chains of the marsupial Tammar Wallaby, *Macropus eugenii*. Eur J Biochem 248:864–871

Holland RA, Gooley AA, Hope RM (1998) Embryonic globins of the marsupial the Tammar Wallaby (*Macropus eugenii*): bird like and mammal like. Clin Exp Pharmacol Physiol 25:740–744

Hosbach HA, Wyler T, Weber R (1983) The *Xenopus laevis* globin gene family: chromosomal arrangement and gene structure. Cell 32:45–53

Hughes JR, Cheng JF, Ventress N et al (2005) Annotation of *cis*-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences. Proc Natl Acad Sci USA 102:9830–9835

Jeffreys AJ, Wilson V, Wood D et al (1980) Linkage of adult alpha- and beta-globin genes in *X. laevis* and gene duplication by tetraploidization. Cell 21:555–564

Opazo JC, Hoffmann FG, Storz JF (2008) Genomic evidence for independent origins of beta-like globin genes in monotremes and therian mammals. Proc Natl Acad Sci USA 105:1590–1595

Opazo JC, Sloan AM, Campbell KL et al (2009) Origin and ascendancy of a chimeric fusion gene: the beta/delta-globin gene of paenungulate mammals. Mol Biol Evol 26:1469–1478

Patel VS, Cooper SJ, Deakin JE et al (2008) Platypus globin genes and flanking loci suggest a new insertional model for beta-globin evolution in birds and mammals. BMC Biol 6:34

Pisano E, Cocca E, Mazzei F et al (2003) Mapping of alpha- and beta-globin genes on Antarctic fish chromosomes by fluorescence in-situ hybridization. Chromosome Res 11:633–640

Washington University Genome Sequencing Center (2010) Internet reference. Retrieved from http://genome.wustl.edu/tools/software/overgo.cgi Accessed 15 April 2010

Wheeler D, Hope R, Cooper SB et al (2001) An orphaned mammalian beta-globin gene of ancient evolutionary origin. Proc Natl Acad Sci USA 98:1101–1106

Wheeler D, Hope RM, Cooper SJ et al (2004) Linkage of the beta-like omega-globin gene to alpha-like globin genes in an Australian marsupial supports the chromosome duplication model for separation of globin gene clusters. J Mol Evol 58:642–652

Yeh RF, Lim LP, Burge CB (2001) Computational inference of homologous gene structures in the human genome. Genome Res 11:803–816

# CHAPTER 5: *Cis*-regulatory regions of the platypus α- and β-globin loci and their evolution in mammals

***Authors:***

**Patel, V.S.**, Dorman, C., King, D.C., Deakin, J.E., Graves, J.A.M, Hardison, R.C.

***Current status of paper (circle as appropriate):***

Planned/ In Preparation/ Submitted/ Under Revision/ Accepted/ Published

***Date paper accepted for publication or anticipated date of publication:***

2011

***Name of journal/book:***

*Molecular Biology and Evolution*

***Extent to which research is your own:***

I conceived this research and performed all computational analyses and experiments, except for the prediction of putative HS of platypus β-globin cluster that was performed by author DK and transfection of cloned constructs into K562 cells that was performed by CD. I also analysed the data.

***Your contribution to writing the paper:***

I drafted the manuscript and prepared figures.

***Comments:***

This project was done in collaboration with colleagues from the Pennsylvania State University (USA). During my visit at Prof. Ross Hardison's laboratory in the U.S.A, he provided valuable help and advice about characterising the regulatory regions of the platypus globin clusters. I was the lead researcher in performing bioinformatics analysis and cloning the test and neutral regions into the MCSγluc vector in our laboratory, after which the samples were sent back to U.S.A for transfection into K562 cells (as we did not have proper equipment and knowledge in our laboratory). I then analysed the data sent to me.

# *Cis*-regulatory regions of the platypus α- and β-globin loci and their evolution in mammals

**Patel, V.S.**[1*], Dorman, C.[2], King, D.C.[2], Deakin, J.E.[1], Graves, J.A.M.[1], Hardison, R.C.[2]

[1] Research School of Biology, The Australian National University, Canberra, ACT 0200, Australia

[2] Center for Comparative Genomics and Bioinformatics, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

* Corresponding author: Vidushi Patel
Research School of Biology
The Australian National University
Canberra ACT 0200, Australia
Tel: +612 61252371; Fax: +612 61258525
Email: vidushi.patel@anu.edu.au

Key words: Evolution, globin, LCR, MRE, platypus, regulation

## Abstract

The high-level expression of genes inside the α- and β-globin loci is controlled by *cis*-regulatory regions containing many DNase I hypersensitive sites (HSs) that bind *trans*-acting factors. These regulatory regions have been extensively studied in eutherian mammals and birds, and comparisons of their sequence and function have been made even with those of fish. However, it will be important to fill the phylogenetic gap with information about globin regulatory regions in monotreme mammals that represent the most basal clade in mammals. In order to fill this knowledge gap, we used a combination of computational approaches to predict an orthologous major regulatory element (MRE of size 389 bp) of the platypus α-globin locus that has a pattern of transcription factor binding sites similar to that of therian mammals. We tested its enhancing activity in the transient transfected K562 erythroid cells, which showed an almost ten-fold increase in activity. We also searched near the platypus β-globin locus for orthologues of the human locus control regions (LCR) and the chicken $\beta^A/\varepsilon$ enhancer, responsible for regulation of the β-globin locus, but failed to find any region conserved between them. We then predicted and tested nine putative sites containing a cluster of binding sites for GATA1, EKLF, AP1/NF-E2 and YY1 that are involved in globin regulation in eutherian mammals. However, the enhancer assay showed only one of these sites to have a significant (two-fold) increase in activity consistent with a role in regulation of the platypus β-globin locus. Our results suggest that the regulatory regions of the monotreme, therian and bird β-globin clusters have either diverged significantly or evolved independently along with their genes. Alternatively, these regions may have evolved by rapid turnover of transcription factor binding sites from a common ancestral regulatory region.

## Introduction

The regulatory regions of the amniote α- and β-globin loci have been used as models for defining the properties of both local and distant non-coding regulatory elements. Despite α- and β-globin clusters being located on different chromosomes, their *cis*-regulatory elements, along with proximal promoters, coordinate the high-level expression and regulation of genes contained within these clusters in erythroid-specific cells and different stages of development. This is critical in producing balanced concentrations of α- and β-globin chains to form a functional $\alpha_2\beta_2$ haemoglobin molecule needed for oxygen transportation and survival.

In therian mammals (marsupials and eutherians), both α- and β-globin clusters possess distinct dominant *cis*-regulatory elements, located further upstream from the structural genes, that can activate the entire cluster and lead to high-level expression of each globin gene. These are known as major regulatory elements (MRE, also known as HS-40 in humans) for the α-globin locus, and the locus control region (LCR) for the β-globin locus. Both regions contain multiple conserved binding sites for *trans*-acting factors (activators and repressors) that confer high-level, tissue-specific, copy number-dependent, position-independent expression of associated globin gene in transgenic mice (Grosveld et al., 1987a, Higgs et al., 1990).

The structure, size and location of α-globin MRE and β-globin LCR are different. The MRE is about 400 bp long and contains a single DNase I hypersensitive site (HS) that is located about 20-60 kb upstream of the α-globin locus. The LCR comprises five HSs (henceforth, referred to as HS1-5), each containing a core sequence of ~250 bp and totalling 20-25 kb in length, and is located 6-20 kb upstream of the β-globin locus (reviewed in Hardison, 1998, Stamatoyannopoulos, 2005). The LCR is also involved in opening the chromatin over a discrete locus in erythroid cells, whereas MRE does not possess this role as it is located in constitutively open and active chromatin in all cell types (reviewed in Hardison et al., 1997, Higgs et al., 2008).

How these regulatory regions evolved is not well understood and their origins are obscure. The α-β globin clusters in fish and frog must represent the ancestral stage, which remains in amniotes as the α-globin cluster (Flint et al., 2001, Gillemans et al., 2003, Hughes et al., 2005, Hardison, 2008, Patel et al., 2008). The MRE of the human

(*Homo sapiens*) α-globin cluster, along with the α-globin cluster of other eutherian mammals, are highly conserved (in terms of distribution of transcription factor binding sites) with MRE sequences in the α-globin cluster of a dasyurid marsupial (the stripe-faced dunnart *Sminthopsis macroura*), chicken (*Gallus gallus*), pufferfish (*Spheroides nephelus*) and medeka (*Oryzias latipes*), reflecting the ancient origin of the cluster (Flint et al., 2001, De Leo et al., 2005, Maruyama et al., 2007).

The amniote β-globin cluster, however, evolved recently, just prior to the reptilian and mammalian divergence that occurred 315 million years ago (MYA), when a copy of the β-globin gene transposed from the ancient α-β globin cluster and inserted into a sea of olfactory receptor genes, followed by lineage-specific duplication of this β-globin gene (Gillemans et al., 2003, Hardison, 2005, Hardison, 2008, Hoffmann et al., 2008, Patel et al., 2008). Therefore, it is not surprising that its regulatory regions have no similarities to the amniote α-globin cluster or fish and frog α-β globin clusters. Given that transposition separated the α- and β-globin clusters in amniotes, it is interesting to consider how this event affected the regulatory apparatus of these globin clusters, in particular the β-globin cluster that resides in a cluster of olfactory genes that are constantly active in olfactory epithelium cells rather than erythroid-specific cells (Bulger et al., 1999). How did the regulatory elements evolve to adapt to the separation of α- and β-globin clusters in amniotes to maintain their coordination, developmental regulation and balanced production?

Perhaps as a consequence of this recent relocation, the LCR of therian mammals shows no conservation with the regulatory region of the chicken β-globin cluster (Reitman and Felsenfeld, 1990, Reitman et al., 1993a). Although chickens have four HSs located upstream of the β-globin locus, these sites do not act as a LCR to enhance the high-level, copy number-dependent, position-independent expression of associated globin gene in transgenic mice (Reitman et al., 1995), nor do they show any sequence similarities to human HS1-5 (Reitman et al., 1993a). Instead, there is a strong enhancer located within the chicken β-globin cluster called the $\beta^A/\varepsilon$ enhancer (Choi and Engel, 1986, Hesse et al., 1986); however, it cannot sufficiently open chromatin, and requires the cooperation of four upstream HSs and other gene promoters (Reitman et al., 1993b).

In this report, we use comparisons between upstream sequences in the duck-billed platypus (*Ornithorhynchus anatinus*), a monotreme mammal, to understand the evolution of MRE and LCR in amniotes. Monotremes are basal in the mammalian phylogeny, having diverged from therian mammals about 166 MYA (Bininda-Emonds et al., 2007). Platypus MRE and LCR have not yet been characterised, although claims have been made for isolating HS2 and HS3 of β-globin LCR in the platypus that shows high sequence similarities to therian HS2 and HS3, respectively (De Leo et al., 2005).

The main aim of this study was to identify MRE and HS1-5 (that represent the LCR) in the sequenced bacterial artificial chromosome (BAC) clones that we had previously characterised and shown to contain α- and β-globin clusters (Patel et al., 2008), in order to study MRE and LCR evolution in mammals. We employed several different bioinformatics strategies to predict these sites, and then tested their activities using enhancer assays by transient transfection into K562 erythroid cells. Our results show that the platypus has a very active orthologous MRE upstream of their α-globin cluster that has a conserved pattern of transcription factor binding sites as therian mammals. However, we failed to find previously reported HS2 and HS3 sites in the platypus genome, nor did we detect any other orthologue of therian HS1-5 sites or chicken $\beta^A/\varepsilon$ enhancer near the platypus β-globin locus. This suggests that platypus sequences may have diverged significantly from those of therian mammals, or that they may possess a unique regulatory region, which may have evolved independently along with their β-like globin genes, or by rapid turnover of transcription factor binding sites, thus showing no conservation at sequence level with therian mammals and birds.

## Materials and methods:

### Data mining and multiple alignment searches for MRE

From the UCSC genome browser (http://genome.ucsc.edu/cgi-bin/hgBlat) (Kent et al., 2002), the genomic region between exons 5 and 6 (including the intermediate intron 5 where MRE is located) of *C16orf35* were extracted for human (hg18 Mar.2006; chr16: 102615-107388), mouse (mm9 Jul.2009; chr11: 32150119-32153963), cow (bosTau4 Oct.2007; ch25: 621310-626170), opossum (monDom5 Oct 2006; chr6: 149327992-149331604) and platypus (BAC clone Oa_Bb2L7 [Genbank accession number:

AC195438]; 118533-121541) and aligned in ClustalW (http://align.genome.jp) (Kanehisa et al., 2006). No alignment was obtained using default settings, but relaxing the stringency of gap penalty to 'gap opening'=5 and 'extension'=2 allowed the sequences to align. The sequences were also aligned in phylogenetic footprinting web-based programs, Multipipmaker (Schwartz et al., 2000) and Mulan (Ovcharenko et al., 2005) using default settings.

**Extraction of short candidate sequences to search for MRE**

Using a short sequence (~400 bp) of the human MRE (hg18 Mar.2006; chr16: 103501-103920), pairwise (local) alignments were performed in ClustalW2 v.2.0.8 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) (Larkin et al., 2007) with each of the above sequences to extract the best aligned region. These candidate regions were short (350-500 bp), representing the orthologous MRE of mouse, cow, opossum and platypus (sequences shown in Supplementary File 1), and were used for multiple alignment in ClustalW2 v.2.0.8 (Larkin et al., 2007) with 'gap opening'=10 and 'extension'=5. The output alignment was further improved manually by either introducing or deleting a gap.

**Database Search for previously isolated HS2 and HS3**

Many different database searches were performed to find the best possible match for previously isolated platypus HS2 [AY692459] and HS3 [AY692460]. These included a BlastN search against the whole database, trace archive of platypus whole genome shotgun (WGS) sequences and other short reads (http://blast.ncbi.nlm.nih.gov/Blast.cgi) (Sayers et al., 2010) and BLAT search against the platypus genome in another genome browser UCSC (Kent et al., 2002). The query sequences were also aligned with a previously characterised platypus BAC clone Oa_Bb484F22 [AC192436] containing the β-globin cluster using Bl2seq (Sayers et al., 2010).

**PCR amplification of HS2 and HS3 from the platypus genomic DNA**

Primers were designed from platypus HS2 [AY692459] and HS3 [AY692460] using Primer3 v.0.4.0 (Rozen and Skaletsky, 2000) and synthesized by GeneWorks Pty Ltd

(Adelaide, Aust). Details of the primers, sizes of amplified products from gDNA and annealing temperature used for PCR reactions are listed in Table 1.

**Table 1:** List of primers, their annealing temperatures for PCR reactions and predicted product sizes of amplification from genomic platypus DNA.

| Primers | Sequences (forward primer on top panel, reverse primer on bottom panel) | Annealing temperature | Product size (bp) |
|---|---|---|---|
| DeLeo_HS2 | CAGCAGTGCTGAGTCATGCTGAGTCATGC<br>ATATTCTAGGTTATGTCACACACACGCCA | 55 | ~200 |
| DeLeo_HS3 | TTTGACTCAGCAAACACAAGACC<br>AGAGTCACAGACTCCACCCTGAGC | 60 | ~200 |
| Own_HS2_1 | GCCAGATGTCAGGAGTGATAGT<br>CCTGGATTAAGTTATTTCCCTTC | 55, 58.1, 61 | 126 |
| Own_HS2_2 | CAGTGCTGAGTCATGCTGAGTC<br>AGGTTATGTCACACAGCAGGC | 55, 58.1, 61 | 198 |
| Own_HS3_1 | CTTTTGAGGCTGCTTTGGC<br>CCCCCTGGTCATCTATCGTG | 55, 58.1, 61 | 126 |
| Own_HS3_2 | CTAGGCCCCCTTTTGAGG<br>ACAGACTCCACCCTGAGCAA | 55, 58.1, 61 | 163 |
| MRE1 | GCACGCGTGCTGCTTCCTTTCCGTTC<br>TTGCGGCCGCATAGGACACGCTGCTTCTGG | 66 | 389 |
| MRE2 | GCACGCGTATGCCAAGAACTGGGGTAGA<br>TTGCGGCCGCACCCAAGTCCATCTGTCTCC | 66 | 821 |
| pHS#1 | GCACGCGTAAACTCATCTTCCACCGCAAA<br>TTGCGGCCGCGGAGTGAAGTGAGTGCCTAAATG | 64 | 848 |
| pHS#2 | GCACGCGTCCTCTCCCTCGTTATGGCTCT<br>TTGCGGCCGCGCTTGGTACATGGCCTAGTGG | 64 | 931 |
| pHS#3 | GCACGCGTTGATCTCCTGGCCTCCTTTC<br>TTGCGGCCGCTGCCACAGAGCTTCCAATCA | 64 | 750 |
| pHS#4 | GCACGCGTTGAGGAACATTTATCGCCTTG<br>TTGCGGCCGCGAGGATGGAGGTGTTGTCTGG | 64 | 835 |
| pHS#5 | GCACGCGTTGAAATGGGAAGCACCCAAG<br>TTGCGGCCGCAATACTGCTTCTCAAGCACTTGGTA | 65 | 655 |
| pHS#6 | GCACGCGTAGGTCCCACATGGTCACAGC<br>TTGCGGCCGCGCTCCTTGGTCATGTGCTTC | 64 | 660 |
| pHS#7 | GCACGCGTGGCCTCGCTTGAACTATGCT<br>TTGCGGCCGCCTCCGATCCTCTGGGTCAGT | 64 | 901 |
| pHS#8 | GCACGCGTGCCTTTACTCCCACCCTCCA<br>TTGCGGCCGCGGCTTGATGGTGATGGCACT | 64 | 680 |
| pHS#9 | GCACGCGTGAGAGATGGAGAACCAAGGACA<br>TTGCGGCCGCTCTTTACACCGCCCAGCTTC | 64 | 655 |
| N#1 | GCACGCGTCCCTCCAATCCCTCCCTGT<br>TTGCGGCCGCCGACTCCTCCAGCCACATC | 66 | 443 |
| N#2 | GCACGCGTGGCACTGGATCACTCTGTGG<br>TTGCGGCCGCTAGCCACCTCTGCCCACTTT | 65 | 908 |
| N#3 | GCACGCGTCACCTCCTTCTTTCATTCTGTCC<br>TTGCGGCCGCCCTTAGTTCCCTGTCCCTTCTG | 65 | 694 |
| N#4 | GCACGCGTCGGGCAAGAGAAGGAGAATG<br>TTGCGGCCGCTGGATGAACCGAGCAGAGTG | 65 | 696 |
| N#5 | GCACGCGTCCTTCACCTTCTCCTCCATCA<br>TTGCGGCCGCTGCTTCACAATTCGCCTTCA | 65 | 669 |
| MCSγluc | GCAAAATAGGCTGTCCCCAGTGCAAGTGCA<br>AGCGGTTCCATCTTCCAGCGGAT | Sequencing primers | |

A standard PCR reaction was performed in a final volume of 25µl, with 100ng platypus female gDNA, 1x Green GoTaq® Buffer (Promega Corporation, Madison, Wisconsin, USA), 0.2mM dNTPs, 1.25U GoTaq® DNA Polymerase (Promega Corporation, Madison, Wisconsin, USA) and 1µM each of forward and reverse primers. PCR cycling conditions were: 94°C for 2 mins, then 35 cycles of 94°C for 30 secs, 56-64°C for 30 secs, 72°C for 1 min, followed by a final extension of 72°C for 10 mins. The amplified products were run on 1% TAE agarose gel stained with SYBR® Safe (Life Technologies Corporation, California, USA). Amplified DNA from the gel was extracted using either QIAquick® gel extraction kit (Qiagen Sciences, Germantown, MD, USA) or PureLink™ PCR purification kit (Invitrogen, Carlsbad, USA) following the manufacturers' protocols. The purified PCR products were then directly sequenced at the Australian Genomic Research Facility (Brisbane, Australia).

Attempts to re-isolate HS2 and HS3 from the platypus gDNA including those from another dasyurid marsupial, the fat-tailed dunnart *Sminthopsis crassicaudata,* and mouse *Mus musculus* (positive control) was also performed following methods as described by De Leo et al. (2005). Although De Leo and colleagues did not specify which brand of DNA polymerase they used in their experiment, we used GoTaq® DNA Polymerase (Promega Corporation, Madison, Wisconsin, USA) for amplification of template DNAs.

**Predicting HSs in the platypus BAC clone containing the β-globin cluster**

Human HS1-5, representing the β-globin LCR, were extracted from the β-globin locus [NG_000007]: HS5 (Coordinates: 6245-6302), HS4 (9648-9713), HS3 (12459-13097), HS2 (16671-17058) and HS1 (21481-21766). Each of these sequences were preliminarily aligned pairwise with the sequenced and characterised BAC clone Oa_Bb484F22 [AC192436] using ClustalW2 v.2.0.8 (Larkin et al., 2007). Even though no significant best-scoring alignment or conserved pattern of binding sites could be seen between human and platypus, the aligned HS3 region was analysed further. This region from platypus BAC clone Oa_Bb484F22 (coordinates: 46315-46585 on reverse complemented strand) was extracted and then multi-aligned with mouse HS3 (mm9 Jul.2009; chr7: 111014292-111014570), cow HS3 (bosTau4 Oct.2007; ch15: 47786477-47786746), dunnart HS3 [AC148753: 18865-19206] and human HS3 (hg18 Mar.2006; chr11: 5262458-5262745) using ClustalW2 v.2.0.8 (Larkin et al., 2007). The

whole method was repeated using stripe-faced dunnart HS3 and chicken $\beta^A/\varepsilon$ enhancer [M12890]. The results were, at best, short, local alignment blocks that were prone to exclude a lot of the platypus sequence, due to lack of conservation over the phylogenetic distances between the species.

A more direct and inclusive approach was then used to predict short candidate regions spanning 500-1000 bp in the platypus Oa_Bb484F22 clone itself, that contained putative binding motif sequences for transcription factors relevant to globin regulation, the results of which could be used to determine conservation. Using nucleotide ambiguity code rules IUPAC (International Union of Pure and Applied Chemistry) consensus binding sites for GATA1, EKLF, AP1/NF-E2 and YY1 (Table 2) were first matched with the platypus sequences, and cases in which these sequences were located in close proximity (spanning 500-1000bp), defined candidate 'clusters' (software available upon request).

**Table 2:** Consensus binding sequences for transcription factors (except for MZF1) implicated in MRE and LCR functions (Hardison et al., 1997 and references therein). MZF1 may regulate tissue-specific expression of α-globin genes along with other transcription factors.

| Motif | Transcription Factor | Expression | Consensus Sequence | Other Relatives |
|---|---|---|---|---|
| **MARE** | AP1/NF-E2 | Ubiquitous | TGASTCA | LCRF1/Nrf1/Bach1 |
| **GATA** | GATA1 | Erythroid | wGATAr | GATA 2, 3, 4, 5, 6 |
| **CACBP** | EKLF | Erythroid | CCnCnCCCn | Sp1, TEF2/BKLF |
| **YY1 site** | YY1 | Ubiquitous | vdCCATnwy | |
| **MZF1 site** | MZF1 | Myeloid | AGTGGGGA<br>CGGGnGAGGGGGAA | |

The sequences of the clusters were preliminarily aligned to the genomic sequences of human, mouse, and dunnart HS3 regions, using MULTIZ (Blanchette et al., 2004) to eliminate false negative predictions. The alignments were then searched for sequences matching the consensus binding sites for GATA1, EKLF, AP1/NF-E2 and YY1 (discarding the indel characters and scanning sequences individually when run on alignments, projecting the position of matches onto a reference sequence). This makes it possible to identify motif-matching sequences in multiple alignments without requiring complete conservation of the motif or complete accuracy of the alignment procedure. In order to determine whether the patterns for transcription factor binding sites were

conserved, multi-alignment with HS3 of human, mouse, cow and dunnart was performed in ClustalW2 v.2.0.10 (Larkin et al., 2007) using 'gap opening'=10 and 'extension'=5. The software for doing this analysis has been implemented as a server at http://cladimo.bx.psu.edu/form/.

**Prediction of other transcription factor binding sites in candidate regions**

Transcription factors were searched in the putative MRE and pHS#7 using TFSEARCH v.1.3 (Akiyama, Unpublished), which searches highly correlated sequence fragments using position-weighted matrices of experimentally proven transcription factor binding sites from the TRANSFAC database (Heinemeyer et al., 1998).

**Designing neutral fragments for the enhancer assay**

Using the platypus genome sequences (ornAna1 Mar.2007) from the UCSC genome browser random non-coding regions from various chromosomes were selected as neutrals (no expected activity) for the enhancer assay. The regions were selected on the basis that they contained comparable length and %GC content to some of the test regions (putative MRE and pHS#1-9). The locations of the regions in the platypus genome are: N#1 (chr1: 38812866-38813308), N#2 (chr3: 26662090-26664089), N#3 (chr1: 13578039-13578732), N#4 (Ultra336: 1518749-1519444), N#5 (chr5: 19109920-19110588).

**Cloning test and neutral regions into MCSγluc plasmid**

There were a total of 16 platypus-specific fragments (2 test for MRE, 9 test for pHS#1-9 and 5 for neutrals) to clone into the firefly luciferase reporter plasmid MCSγluc (Elnitski et al., 2001) containing the human $^{A}$γ-globin (*HBG1pr*) gene promoter (Wang et al., 2006).

PCR was performed as above except for some modifications. Restriction sites for *Mlu* I (GCACGCGT) and *Not* I (TTGCGGCCGC) were added at the 5' ends of the forward and reverse primers, respectively (refer to Table 1 for primer details). For the PCR reaction, instead of using GoTaq® DNA Polymerase, 0.5x Advantage® 2 DNA Polymerase and 1x Advantage® 2 Buffer (Clontech Laboratories, Inc., Mountain View,

CA, USA) were used. The PCR reaction was also increased ten-fold to result in a highly concentrated end product. DNA double digest of the PCR products (~200ng) and vector (450ng) were performed in a total volume of 10µl with 0.1µl of 10x BSA Buffer, 1µl of 10x Buffer D, 0.5µl of *Not* I restriction enzyme and 0.5µl of *Mlu* I restriction enzyme (all from Promega Corporation, Madison, Wisconsin, USA) and incubated at 37°C for two hours. The reaction was stopped at 65°C for 20 mins. DNA was ligated in a final volume of 10µl with the vector to insert ratio of either 1:2, 1:3 or 1:4 and with 1x T4 ligase buffer and 3U of T4 DNA Ligase (Promega Corporation, Madison, Wisconsin, USA). The samples were incubated at 15°C for two hours. The cloned plasmid was transformed into TOPO One Short® TOP10 competent cells using TOPO TA cloning® kit protocol (Invitrogen, Carlsbad, USA) and the resulting plasmids were purified according to Wizard® Plus SV Minipreps DNA Purification System (Promega Corporation, Madison, Wisconsin, USA). To confirm if the insert contained the region of interest, DNA double digest was performed (as above) and the insert was directly sequenced at the Australian Genome Research Facility (Brisbane, Australia) using plasmid-specific (MCSγluc) primers shown in Table 1.

**Enhancer assay by transient transfection**

The enhancer assay was carried out according to Cheng et al. (2009) with the following modifications. The cloned plasmid DNAs, as well as three previously tested mouse-specific neutrals (GHN78, GHN213 and GHN240) and two positive controls (GHP88 and GHP181) from our last experiment, were transiently transfected into K562 cells in a 96 well plate (eight wells per test construct). In each well, 14µg of cloned plasmid containing firefly luciferase reporter was mixed with 0.00035µg of cotransfection control plasmid expressing Renilla luciferase in OptiMEM medium (Invitrogen, Carlsbad, USA), 0.21µl Lipofectamine LTX and 14µl of PLUS Reagent (Invitrogen, Carlsbad, USA). The K562 cells were plated at 2.8 x $10^4$ cells per well in 200µl. Each plasmid was transfected eight times and tested in four separate experiments to get an average. Luciferase activities were recorded after 48 hours.

## Results

Many different approaches were utilised to search for regulatory regions of the platypus α- and β-globin clusters. To identify α-globin MRE in the platypus, long sequences (between exon 5 and 6 of *C16orf35*) and short candidate sequences (containing MRE) were aligned in multiple sequences using different phylogenetic footprinting programs. However, identifying the platypus β-globin HSs were not as straight forward as identifying the α-globin MRE. Here we describe a series of methods that we used to identify the regulatory regions of the platypus β-globin cluster and their outcomes. Predicted regulatory regions were experimentally tested for their ability to enhance expression of reporter gene in transiently transfected K562 cells.

### Predictions of the platypus α-globin MRE

The first approach used to identify MRE in the platypus BAC clone OaBb2L7 was to align the region between exons 5 and 6 of *C16orf35* (where MRE resides in therian mammals) in multiple species. This should allow programs to anchor (i.e. align) both exons and reveal conservation in the intermediate intron that would help identify MRE in the platypus sequence.

Many well-known programs, for example, ClustalW2 (Kanehisa et al., 2006), Multipipmaker (Schwartz et al., 2000) and Mulan (Ovcharenko et al., 2005) that reveal conserved regions or phylogenetic footprints in non-coding sequences, were used for comparative analysis between human, mouse, cow, opossum and platypus. However, none of these programs detected any conservation between the platypus intron and the homologous region in the other species. Using ClustalW2 with the stringent default settings, an alignment of exon 5 but not exon 6 was observed. After relaxing the parameters (see Materials and methods), it was possible to anchor exons 5 and 6 on both sides, but the alignment between them showed very low sequence similarities between different species.

Using Multipipmaker, the graphical visualisation of multiple alignment showed no signals between exons 5 and 6 of *C16orf35* when the platypus sequence was aligned to the homologues in therians. Lastly, using Mulan, which aligns multiple finished or draft secondary sequences using local alignment, exons 5 and 6 were anchored nicely and there were some conserved signals seen in the intron in some species (several in human,

mouse and cow and one small signal in opossum) but no signal was detected in the platypus sequence.

Therefore, all the well-known approaches failed to identify a conserved MRE in the platypus sequence, unlike placental mammals and marsupials, in which a MRE was obvious. While the alignment methods could have missed very short regions of homology, it appears that this non-coding sequence has diverged so much, over the long time that platypus and therian mammals were independently evolving, that it can no longer align to the therian homologues as a full intron (anchored in exons).

We considered that non-conserved portions of this intron that are not involved in gene regulation could be adding so much "noise" that they preclude discovery of alignments in the functional (and hence probably most conserved) portion of the intron. Thus we refined our search by subdividing the sequence into short candidate regions to focus only on the MRE region of size 350-500 bp. For this, we first started with the human MRE and performed many pairwise local alignments with each of the above sequences in mouse, cow, opossum and platypus. This resulted in a short region of best alignment (candidate region), which was then extracted from all species, followed by a multiple alignment of these candidate sequences in ClustalW2 (Larkin et al., 2007). The resultant alignment showed four conserved motifs; two that match the MAf Recognition Elements (MARE), to which AP1/NE-F2 transcription factors bind, and two that match the consensus binding site for GATA1 and its relatives (referred to subsequently as a GATA site) (Supplementary File 2). Previous studies of HSs in therian mammals identified a phylogenetic footprint in which four out of six base pairs matched between marsupial and eutherian mammals (De Leo et al., 2005). We also used the same criteria to identify a binding motif that showed at least 66% conserved matches in the consensus binding sites for transcription factors between the platypus and human.

Upon further investigation of the multi-alignment, we found that alignment was not optimal as it failed to align one GATA site properly that was located 5' to the first MARE site and two CACBP sites for binding EKLF in eutherian mammals. We therefore improved the alignment manually by either introducing a gap or deleting a gap. The manual alignment now showed a clear pattern of conservation among five different mammalian species (Figure 1).

The approach of performing many pairwise alignments, extracting short candidate regions that show similarities, doing multiple alignment and then improving the alignment manually, was finally successful in revealing five motifs (GATA, MARE, MARE, GATA, GATA) conserved between the platypus and the therian species, and two additional motifs (GATA, CACBP) conserved from opossum through placental mammals. The order of these motifs has been preserved in therian mammals and the platypus (Figures 1 and 2A).

We also discovered two highly conserved phylogenetic footprints of eight consecutive matches that have not previously been identified; CCCCTCCT is conserved in eutherian mammals only (except for one mis-match in the mouse), and GGGTGGRG is conserved in all five mammals (Figure 1). The CCCCTCCT sequence is also present in opossum and platypus, but has been interrupted by a small block of sequence. Further analysis using TFSEARCH against the TRANSFAC database (Heinemeyer et al., 1998, Akiyama, Unpublished) showed that in eutherians this conserved footprint contains a binding site for myeloid zinc finger protein 1 (MZF1) that regulates haematopoiesis (in particular granulopoiesis) and is involved in oncogenesis (Bavisotto et al., 1991, Hromas et al., 1991, Gaboli et al., 2001). Although MZF1 consensus sequences had been disturbed in the platypus and opossum, TFSEARCH predicted one other binding site for MZF1 in the platypus and four others in the opossum but they were randomly distributed and were not conserved with each other or with eutherian sequences.

The other conserved footprint (GGGTGGRG) is highly conserved in all extant mammals. However, TFSEARCH could not detect any recognisable binding site for any of the transcription factors present in the database in this conserved footprint.

To gain further insight into the platypus MRE, TFSEARCH was used to predict other transcription factor binding sites. Platypus MRE contains many putative binding sites for many other transcription factors such as CREB (camp response element binding protein), MyoD (myoblast determination gene product), SRY (sex-determining regulatory region Y gene product), HLF (hepatic leukemia factor), Pbx1 and CdxA homeobox (Figure 2A). Interestingly, some of these transcription factors (MyoD, HLF and PbX1) are involved in blood cell differentiation and leukaemia and so have a function similar to that of MZF1.

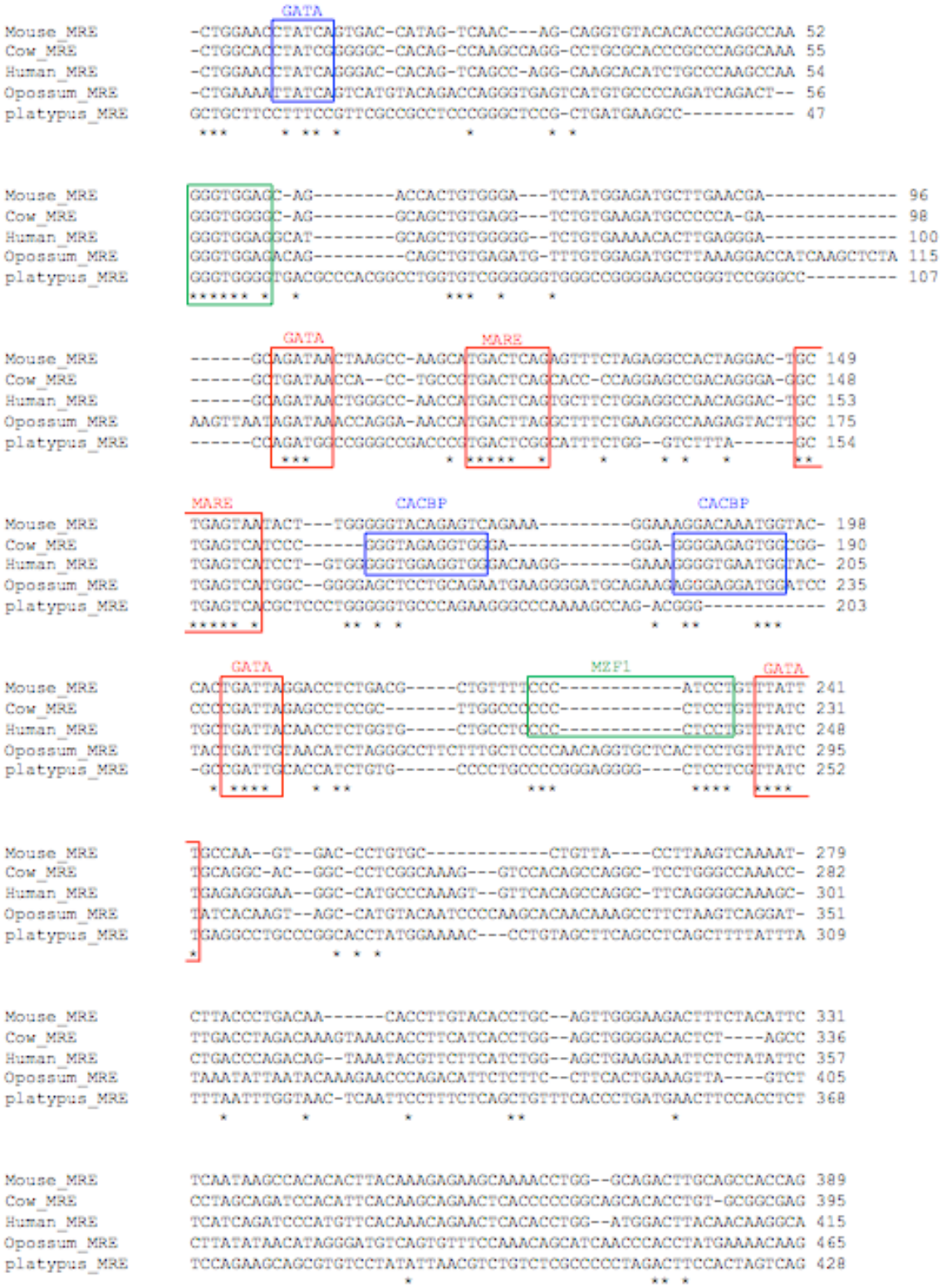
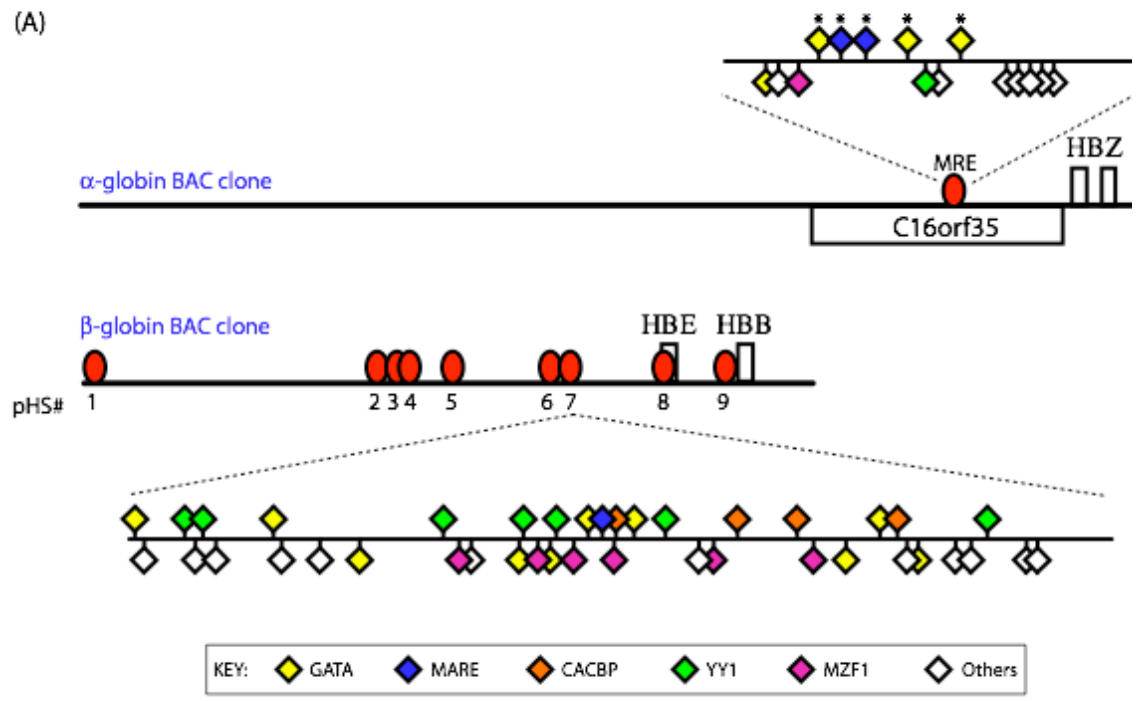



**Figure 1: Improved multi-alignment of orthologous MRE in the mouse, cow, human, opossum and platypus showing conserved pattern of transcription factor binding sites.** A phylogenetic footprint in the platypus was identified in which four out of six base pairs matched with the human based on their divergence time of 166 million years. The *red* boxes show conservation in all five species, *blue* boxes show conservation in some mammals, and *green* boxes show phylogenetic footprint identified in this experiment that were not previously identified.



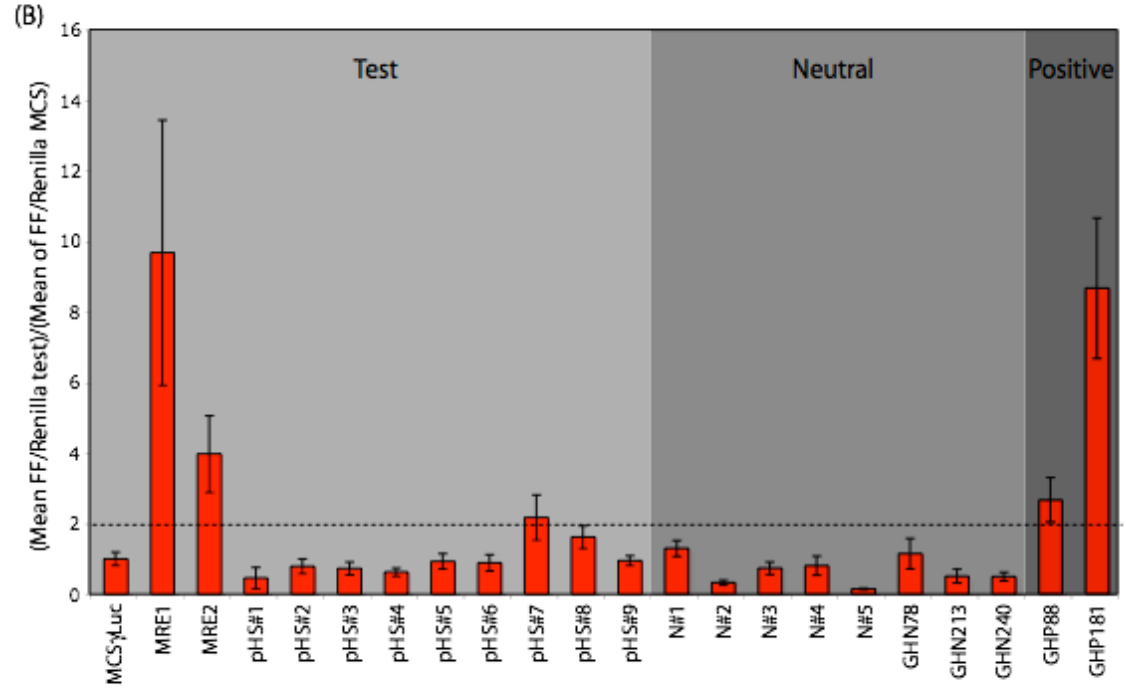




**Figure 2: Locations and validations of the predicted regulatory regions of the platypus α- and β-globin clusters.** (A) Locations of predicted MRE and pHSs in the platypus BAC clones Oa_Bb2L7 and Oa_Bb484F22, respectively. The dashed lines above MRE and below pHS#7 show detailed information about the binding sites of various transcription factors; diagonals above the line summarises our findings (that is, multiple alignment for MRE and prediction of clusters containing globin-specific binding motifs in pHS#7) and diagonals below the line show predictions by TFSEARCH. *Coloured* diagonals (except *purple*) show locations of well-known motifs involved in globin regulation, *purple* diagonals show predicted sites for MZF1 protein,

and *white* diagonals show the presence of predicted binding sites for other transcription factors that have not yet been tested for their occupancies or role in globin synthesis. There are six *white* diagonals below the MRE line, arranged in the order 5'-CREB-MyoD-Pbx1-CdxA-HLF-SRY-3', whereas 12 in the pHS#7 are arranged as 5'-CdxA-Pbx1-CdxA-CdxA-SRY-CdxA-SRY-CdxA-CdxA-AP1-CdxA-SRY-3'. Asteric (*) above diagonals shows high conservation with other therian mammals. (B) Transient enhancer essay of eleven test regions (MRE1-2 and pHS#1-9), five platypus-specific neutrals (N#1-5), three known neutrals (GHN78, GHN213 and GHN240) and two positive controls (GHP88 and GHP181) from our last screen. All ratios have been normalised against the parental reporter gene plasmid (MCSγluc without any insert) and those that have more than two-fold increase in expression are considered as active enhancers or promoters.

---

**Search for HSs in the platypus β-globin cluster**

In order to locate platypus HSs within β-globin cluster, we searched for the HS2 and HS3 regions previously isolated by De Leo et al. (2005) in the sequenced BAC clone Oa_Bb484F22 containing the β-globin cluster, and in two different databases containing platypus genome sequence (assembled whole genome sequences and trace archives). No matches were obtained. Even when we tried to amplify the region from the platypus gDNA using sequences from both De Leo et al. (2005) and our own redesigned PCR primers, no product was amplified, although amplification was seen for mouse, the positive control (Supplementary File 3). These results using two different approaches casts doubt on previous claims to have isolated HS2 and HS3 in the platypus. It is possible that De Leo and colleagues had some contamination with other mammal DNA in their platypus samples.

We then used sequences of human HS1-5, dunnart HS3, and chicken $\beta^A/\varepsilon$ enhancer to search for homologies in the platypus Oa_Bb484F22 clone. Similarities between each of these sequences and the platypus sequence were very low and contained no clear conservation in the region. At most, only one conserved motif could be tentatively detected among them, but the adjacent alignments contained many mis-matches and gaps, and since they did not correspond to the conserved pattern of motifs, we concluded that these regions do not represent a true HS. Therefore, the same approach used for predicting platypus α-globin MRE proved to be unsuccessful in predicting any conserved HS in the platypus β-globin locus. Thus, even if HS exist in the β-globin

locus, it appears that the adjacent sequences have diverged beyond the similarity level necessary for local alignment tools to identify them within the BAC (using sequences of other species as queries).

However, it remains possible that putative binding motif sequences for transcription factors relevant to globin regulation exist close to each other in a regulatory region, and, if conserved, could facilitate a more direct alignment between smaller candidate sequences. We therefore used another approach of predicting short regions spanning 500-1000 bp containing a 'cluster' of motif binding sequences for many known transcription factors (Table 2). Using this approach, we predicted fifteen HS sites across the entire BAC clone that fulfilled these criteria. Nine (referred to hereafter as pHS#1-9) were selected that contained many GATA, CACBP, MARE and YY1 sites (Figure 2A). Their distance from the start codon of the first gene (*HBE*) located inside the platypus β-globin cluster are: pHS#1 (-79 kb), pHS#2 (-40 kb), pHS#3 (-37 kb), pHS#4 (-35 kb), pHS#5 (-28 kb), pHS#6 (-26 kb), pHS#7 (-12 kb), pHS#8 (+0.5 kb), pHS#9 (+9 kb). Negative values indicate that the putative sites are located upstream of *HBE* and positive values indicate that the sites are located downstream of *HBE*.

One of the predicted HSs (pHS#2), located ~40 kb upstream of platypus β-globin cluster, contained a pattern of binding motifs similar to that of therian HS3. A comparison of this predicted region with HS3 of human, mouse, cow and dunnart showed one highly conserved consensus sequence for MARE (no mismatches in the multiple alignment of 8 bp) and three others for GATA sites (4 out of 6 bp) (Figure 3). However, the region contained no conserved CACBP or YY1 sites. All nine predicted pHSs were then tested for regulatory activities using a transient enhancer assay.

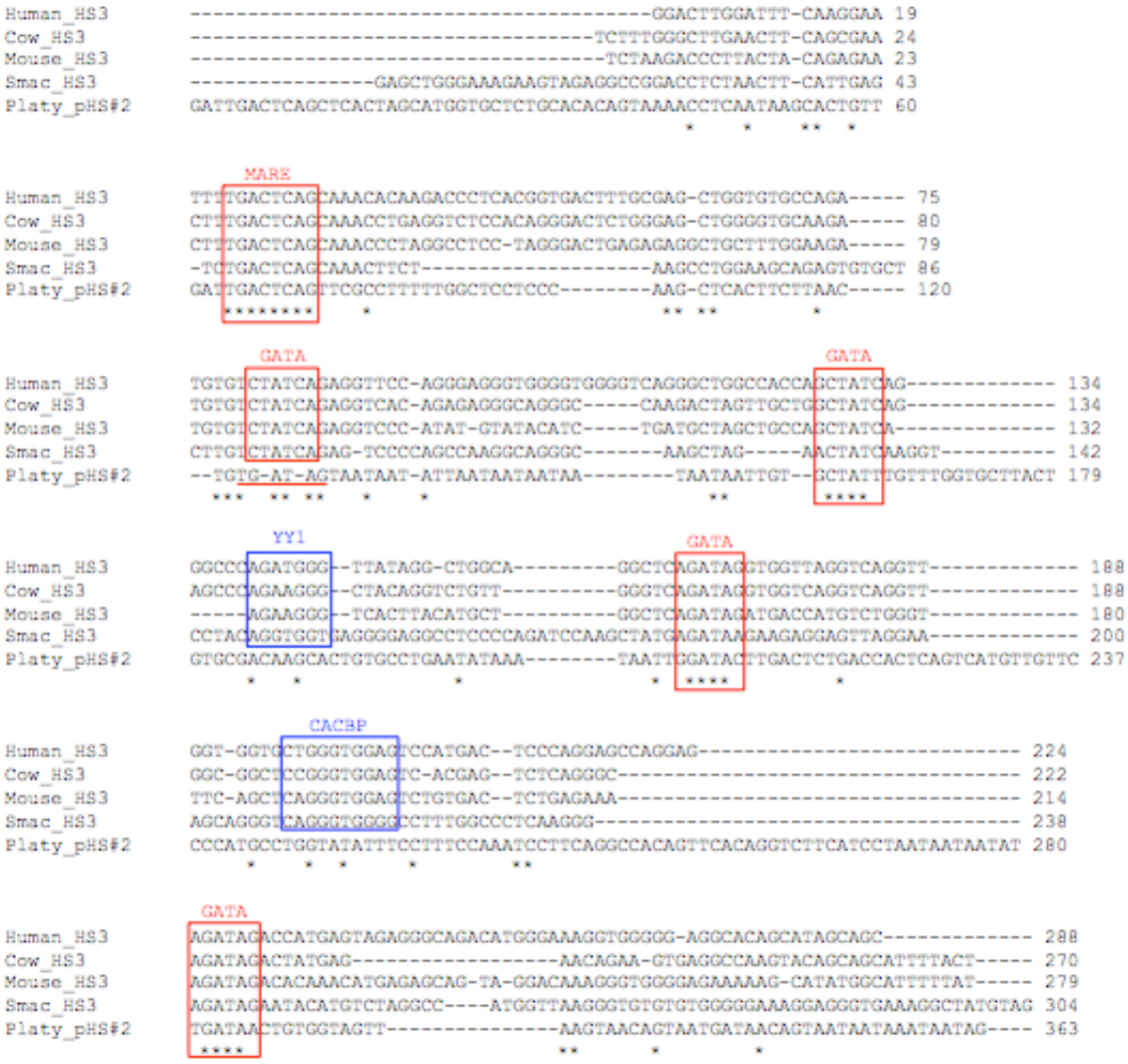



**Figure 3: Improved multi-alignment of platypus pHS#2 with HS3 of human, cow, mouse and stripe-faced dunnart (*Sminthopsis macroura*) showing conserved pattern of transcription factor binding sites.** A phylogenetic footprint in the platypus was identified in which four out of six base pairs matched with the human based on their divergence time of 166 million years. The *red* boxes show conservation in all five species and *blue* boxes show conservation in some mammals. Note: the first GATA site does not align properly with the platypus sequence but there is a GATA site on the opposite strand (underlined).

**Enhancing activities of predicted MRE and pHS regions**

To investigate the extent to which MRE and pHS predicted regions correlate with measurable regulatory activity, we tested 21 constructs for their ability to enhance expression of the luciferase gene from an *HBG1* promoter in transiently transfected K562 cells (Figure 2B). The constructs were two for MRE, nine for pHS, two positive controls, and the rest as neutrals with an expectation of no activity. The constructs that showed at least a two-fold increase in activity compared to that of the parental reporter gene plasmid (MSCγluc) were considered to be active as enhancers (Cheng et al., 2009).

We tested the entire platypus MRE region for enhancing activity using two cloned products: one short fragment (MRE1 of size 389 bp) that contained only the aligned region, and a longer fragment (MRE2 of size 821 bp) that contained an extra 200 bp on both sides of the shorter fragment. This should indicate whether the platypus MRE has maintained a functionally similar role to that of the human and mouse MREs. Both constructs showed extensive increase in expression, by almost ten-fold for MRE1 and four-fold for MRE2 (Figure 2B).

We tested all nine of the β-globin predicted pHSs, finding only one out of nine test constructs (pHS#7) that showed a significant (two-fold) increase in expression (Figure 2B). This region is located about 12 kb upstream of the β-globin cluster and contains 100% identical matches to the consensus sequences of GATA (five present), CACBP (four), MARE (one) and YY1 (seven) sites. Further predictions by TFSEARCH identified five additional matches to a position weight matrix formulation for GATA sites and many other transcription factor binding sites such as CdxA (seven), MZF1 (six), SRY (three), Pbxl (one) and AP1 (one) in this region (Figure 2A). No significant activity was observed for pHS#2, the sequence that showed some similarity to HS3 of therian mammals.

## Discussion

It is often difficult to find conserved regulatory elements in non-coding regions of distantly-related species because sequences have diverged significantly, and signals of conservation are lost. Even when comparing long sequences, alignment programs can at

times overlook very short conserved motifs. Consequently, there is no universally applicable method for predicting conserved regulatory sites.

**Discovery of MRE in the platypus, and its ancient origin**

MRE is a small and powerful region of ~400 bp containing many binding sites for important transcription factors that can activate the entire α-globin cluster and lead to high-level expression of each α-like globin gene. We found that many well-known programs with fixed settings and arbitrary criteria failed to identify the platypus MRE in the fifth intron of *C16orf35*, perhaps due to our use of large divergent genomic sequences. Instead, a simple combination of different programs, along with manual alignment to improve the multi-alignment, was successful in discovering a short 389 bp region in the platypus α-globin cluster that is orthologous to the therian MRE. We showed that this sequence is also active as an enhancer after transfection into a hematopoietic cell line. This sequence conservation and activity make a strong case for the identification of the MRE of the platypus α-globin cluster.

The remainder of the intronic sequence containing the platypus MRE was not alignable with sequence from therian mammals, but the distribution pattern of transcription factor binding sites was conserved among them. Platypus MRE has five highly conserved motifs in the order GATA-MARE-MARE-GATA-GATA, in common with therian mammals, suggesting that these motifs were present in the common ancestor of all mammals. Two additional motifs (GATA and CACBP) are highly conserved between eutherian and marsupial mammals, in the order GATA-GATA-MARE-MARE-CACBP-GATA-GATA. These two motifs are absent from the platypus MRE, implying either that these two sites evolved in the therian ancestor after their divergence from monotremes, or these two sites were also present in the mammalian ancestor but were secondarily lost in the platypus lineage.

Likewise, it has been difficult to identify MRE in more distantly related vertebrates. Flint et al. (2001) found that previously described methods failed to identify MRE of the α-globin clusters in chicken and Southern pufferfish (*Spheroides nephelus*). They also used their own method where they first searched for paired MARE sites (separated by 21 bp) upstream of the chicken and pufferfish α-globin cluster. After finding two paired MARE sites, they extended their search to a neighbouring region to search for

GATA sites. Using this method, Flint and others discovered the MRE in the fifth intron of *C16orf35* in both species, and these sequences also contained a distribution pattern of transcription factor binding sites (GATA-MARE-MARE-GATA) similar to human and mouse MREs. The presence of conserved MRE in the same position in amniotes and teleost fish suggests that MRE has an ancient origin, and arose early in the evolution of the α-globin cluster (Flint et al., 2001). It may have evolved ~450 MYA in the fifth intron of *C16orf35* before the duplication of the ancestral α- and β-globin gene(s), and since then, it has been under strong evolutionary constraint (negative or purifying selection) to maintain similar function in regulating the developmental stage- and tissue-specific expression of α- and β-globin genes in jawed vertebrates and the α-globin genes of amniotes.

**The platypus homolog to therian α-globin MRE is an active enhancer**

The predicted platypus MRE was confirmed to have regulatory activity by an enhancer assay. It produced almost a ten-fold increase in expression, demonstrating that is capable of serving as an enhancer. This strongly suggests that it has a similar function to the human and mouse MRE, activating the entire endogenous α-globin cluster in the platypus. However, its enhancing power was reduced by 2.5-fold in the presence of adjacent sequences (MRE2 in Figure 2B). Perhaps adjacent sequences embed sites for negative regulators (insulators, repressors and/or silencers), which reduces the enhancer activity of MRE in K562 cells. Alternatively, the longer sequence could reduce activity by moving the active enhancer further away from the target promoter.

We discovered two conserved regions that were not previously identified in any eutherian MRE. One has a consensus sequence of GGGTGGRG that does not correspond to any known transcription factor binding site, and could either be an evolutionarily conserved repeat or a novel binding motif that was present in the common mammalian ancestor. The other has a consensus sequence of CCCCTCCT that is recognised by MZF1 transcription factor. MZF1 belongs to a class of SCAN domain containing 13 $C_2H_2$ zinc finger proteins arranged in a bipartite DNA binding domain: fingers 1-4 bind to consensus sequence AGTGGGGA and fingers 5-13 bind to CGGGnGAGGGGGAA (Hromas et al., 1991, Morris et al., 1994). It is preferentially expressed in myeloid progenitor cells, and has a critical function in regulating blood cell proliferation and possibly in acting as a tumor suppressor in myeloid progenitors during

haematopoiesis (Bavisotto et al., 1991, Hromas et al., 1991, Gaboli et al., 2001). Abnormal production of MZF1 can alter its binding properties and regulatory functions, and so could promote many different cancers (Inoue et al., 2005, Peterson et al., 2006, Mudduluru et al., 2010).

It is interesting that there is a highly conserved MZF1 binding site in eutherians (except for mouse, which has a point mutation that obviates recognition of MZF1). This region is also present in the platypus and opossum but has been interrupted by a block of sequences, suggesting that a MZF1 binding site was present in the mammalian ancestor 166 MYA and has been retained as a putative functional site in the eutherian lineage only. Functional experiments such ChIP-chip analysis or mutagenesis experiments in mice could shed light into their roles and importance to globin synthesis.

**Orthologues of therian HS1-5 or chicken $\beta^A/\varepsilon$ enhancer are not present in the platypus β-globin locus**

Therian HS1-5 each function as either an enhancer, insulator or boundary element but all together form a powerful LCR that opens chromatin and leads to a high level, stage-specific and copy number-dependent expression of β-like globin genes in transgenic mice (Grosveld et al., 1987a, Higgs et al., 1990). Chickens also have four HSs upstream of their β-globin cluster and a strong enhancer located within the cluster with similar functions but both regions (as a group) do not act as a LCR because they need contribution from other local promoters to open chromatin (Reitman et al., 1993b). Our results show that the platypus does not have orthologues of therian HS1-5 and chicken $\beta^A/\varepsilon$ enhancer but has a unique pHS#7.

It is evident from our search, by two different methods, that the platypus lacks HS2 and HS3 sequences. Our results are therefore inconsistent with a previous report of their isolation from platypus genomic DNA using conserved primers across eutherians (De Leo et al., 2005). Our failure to find HS sequences in the upstream sequences of the platypus β-globin cluster, and even in the entire genome, suggests possible contamination of platypus DNA samples used by De Leo et al. (2005), perhaps by DNA from a marsupial species, as they show very high conservation with eutherian HS2 and HS3 sequences.

Platypus sequences in the ~120 kb region containing the β-globin locus also show no similarities to therian HS1-5 or to chicken $\beta^A/\varepsilon$ enhancer, at least not at the sequence level. This leaves open the possibility that orthologues of either the therian HS1-5 or bird $\beta^A/\varepsilon$ enhancer may be located elsewhere in the platypus genome, which will require extensive research to find these regions and confirm their function. We hypothesise that the regulatory regions of the platypus β-globin cluster have diverged so significantly that they no longer show similarities to either therian or chicken regulatory regions. Thus platypus may contain a unique set of HSs, enhancers, insulators, silencers and promoters for regulating their β-like globin genes.

Our prediction of nine pHS for the platypus β-globin cluster identified one region (pHS#2) that shared matches to two transcription factor binding site motifs within the therian HS3. However, it did not cause any increase in expression of the luciferase reporter gene, and thus it does not act as an enhancer in this assay. Instead, another region (pHS#7) located approximately 12 kb upstream of the platypus *HBE*, was found to be active, exhibiting a two-fold increase in expression. We hypothesise that pHS#7 is part of the regulatory region for platypus β-like globin genes. By analogy to the locus control region of the homologous genes in eutherian mammals, it may act in concert with other DNA segments to regulate the expression of the β-like globin genes. Several of the HSs of eutherian locus control regions show no independent activity in enhancer assays, and it is possible that some of the platypus predicted HSs with no enhancer activity could play a role modulating the activity of pHS#7. The platypus pHS#7 contains many GATA, MARE, CACBP and YY1 binding motifs, but their arrangement is not similar to those of human HS1-5 or chicken $\beta^A/\varepsilon$ enhancer. These factors may coordinate the regulation of both α- and β-globin genes in the stage- and tissue-specific manner.

Thus, ChIP data and mutagenesis is important for future work. ChIP experiments can help determine whether the predicted binding motifs are occupied by transcription factors, in order to give insight into their functions. This information can then help in mutagenesis or knock-down experiments in transgenic mice where the predicted binding motifs are either mutated or deleted and cloned into the expression vector, similar to the one used in this experiment, and then transformed into transgenic mice. The effects of these mutations or deletions would then be observed for their ability to either up-regulate or down-regulate the expression of the reporter gene, thereby

providing information on the importance of these binding motifs to β-globin synthesis and regulation.

The absence from the platypus genome of observable homologues to either the therian HS1-5 or the bird $\beta^A/\varepsilon$ enhancer makes sense in terms of the origin and expression of genes contained in their β-globin clusters. Platypus, like marsupials, have two β-like globin genes in the order 5'-*HBE*-*HBB*-3', but they are not in a 1:1 orthologous relationship to therian or bird *HBE* and *HBB* globin genes (Opazo et al., 2008b). Phylogenetic analysis showed that both genes grouped closely to *HBB* of another monotreme mammal, the echidna (*Tachyglossus aculeatus*), suggesting an independent duplication of the ancestral β-globin gene, separate to those of therian mammals and birds (Opazo et al., 2008b, Patel et al., 2008). Since the bird and mammalian β-globin clusters have already been shown to have evolved independently from a single β-globin gene in the common ancestor (Czelusniak et al., 1982, Reitman et al., 1993a), these results suggest that there were three independent duplications of an ancestral β-globin gene after transposition, one each in the bird, monotreme and therian lineages (Hoffmann et al., 2010b). Therefore, the number of genes contained in the β-globin cluster, the expression profile and the locations of *cis*- regulatory regions do not correlate across species, as a result of the independent duplications in the separate lineages.

**Independent origins of regulatory regions for the amniote β-globin clusters**

Two models were originally proposed for the evolution of regulatory regions for the amniote β-globin clusters (Hardison, 2008). Both assume that the transposition of the β-globin gene into the olfactory region occurred before reptile and mammalian divergence 315 MYA. The first model hypothesises that the transposed gene carried some regulatory sequences, which underwent considerable divergence, or possibly duplications, after insertion (Figure 4A). The second model hypothesizes that new regulatory regions were formed as a result of the insertion of the transposed β-globin gene (Figure 4B). Our data from a monotreme species neither support nor refute either model.

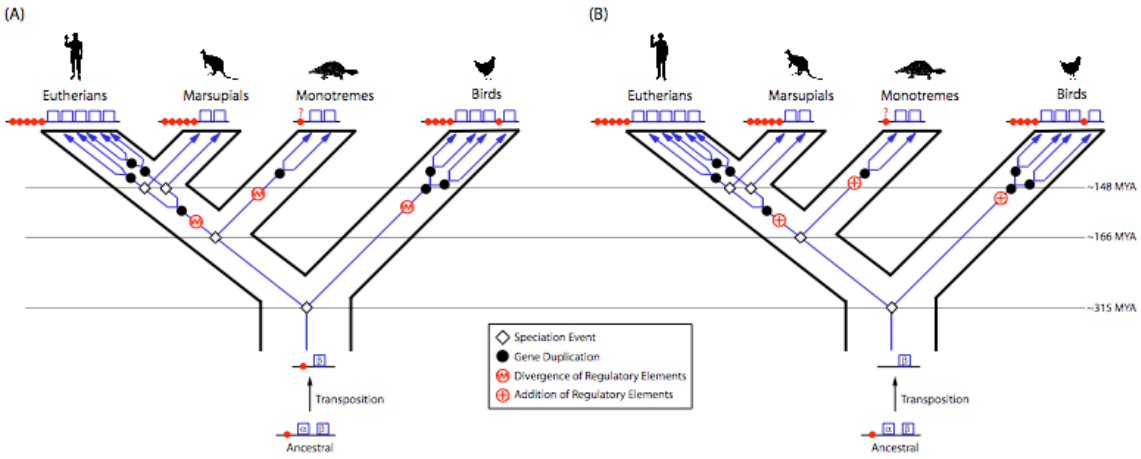



**Figure 4: Two hypothesized models for the evolution of regulatory regions of the amniote β-globin clusters.** (A) According to this model, some regulatory elements from the ancestral α-β globin region were also transposed with the β-globin gene, which after insertion underwent significant divergence followed by duplications in each amniote lineage. (B) According to this model, after the insertion of the transposed β-globin gene, new regulatory elements evolved independently in each of the amniote lineages, followed by further duplications. Both models show three independent events occurring in amniotes.

The first model makes sense of how birds and mammals have maintained stage- specific erythroid expressions but have not maintained sequence and functional similarities. Even though the human LCR has been shown to create stage-specific erythroid expression of human β-globin genes, the developmental stage specificity of the chicken β-like globin genes do not depend on major regulatory elements (both the upstream HSs and intermediate enhancer) but on the gene promoters (Mason et al., 1995). We hypothesise that platypus regulatory sequences have also diverged so significantly that they no longer show any sequence similarities to either chicken or therian *cis*-regulatory elements, and that the ancestral signals for creating stage-specific erythroid expression of globin genes have been lost. However, this suggests three independent divergences of a highly important regulatory region in each of bird, monotreme and therian lineages. This seems less parsimonious, given that most important regions such as regulatory elements evolve under evolutionary constraints (purifying or positive selection) that leads to preservation of transcription factor binding sites over a wide phylogenetic distance (Cheng et al., 2009), as illustrated by α-globin MRE that have maintained similar locations, pattern of binding motifs and function for over 400 MYA.

Why, then, did β-globin regulatory regions diverge so greatly in birds, monotremes and therian mammals that they left behind more differences than similarities? The observation that the same transcription factors (GATA1, AP1-NF-E2, EKLF) bind to bird and mammalian β-globin regulatory regions suggests a rapid 'turnover' of transcription factor binding sites. New motifs have arisen in the vicinity of an old binding site, and their use allowed the original one to diverge and become redundant. This results in new functional binding sites in a slightly different location from the original site. Changes in the surrounding sequences are so extensive that the regulatory regions can no longer be brought into alignment.

## Conclusion

The questions about the evolutionary origins of the regions that regulate the α- and β-globin loci have been difficult to answer without information from all three major mammalian lineages. Our search for these regulatory regions in the platypus has shown that the ancient α-globin regulatory element is conserved in sequence, transcription factor binding motifs and its ability to enhance gene expression amongst vertebrates, reflecting the stability of this regulatory region during vertebrate evolution. Conversely, exhaustive searches for the platypus β-globin regulatory region have failed to uncover a regulatory element conserved either between platypus and therian mammals, or platypus and chicken, indicating either a more rapid (more substitutions) or more extensive (more rearrangements, e.g.) mode of evolution from a common ancestral regulatory region. Alternatively, the regulatory regions of β-globin genes may have independent origins in different amniote lineages.

## Acknowledgements

We thank Swathi A Kumar (Pennsylvania State University, USA) for providing information about cloning inserts into the MCSγluc plasmid, Melanie Edwards (The Australian National University, Australia) for providing help with drawing graphs in excel and Ke-Jun Wei (The Australian National University, Australia) for her help in one of the PCR reactions.

## Supplementary Materials

**Supplementary File 1:** Sequences of MRE in the human, mouse, cow and opossum were extracted from their genome browsers (refer to methods for coordinates). The platypus MRE was extracted from Oa_Bb2L7 [AC195438: coordinates 118533-121541] and platypus pHS#7 was extracted from Oa_Bb484F22 [AC192436: coordinates 66939-67946 on reverse complement strand]. In the platypus pHS#7, underlined letters represent consensus binding sites GATA, CACBP and MARE and YY1 that were predicted.

>Human_MRE
CTGGAACCTATCAGGGACCACAGTCAGCCAGGCAAGCACATCTGCCCAAGCCAAGGGTGGA
GGCATGCAGCTGTGGGGGTCTGTGAAAACACTTGAGGGAGCAGATAACTGGGCCAACCATG
ACTCAGTGCTTCTGGAGGCCAACAGGACTGCTGAGTCATCCTGTGGGGGTGGAGGTGGGAC
AAGGGAAAGGGGTGAATGGTACTGCTGATTACAACCTCTGGTGCTGCCTCCCCCTCCTGTTT
ATCTGAGAGGGAAGGCCATGCCCAAAGTGTTCACAGCCAGGCTTCAGGGGCAAAGCCTGAC
CCAGACAGTAAATACGTTCTTCATCTGGAGCTGAAGAAATTCTCTATATTCTCATCAGATCC
CATGTTCACAAACAGAACTCACACCTGGATGGACTTACAACAAGGCAGCATC

>Mouse_MRE
CTGGAACCTATCAGTGACCATAGTCAACAGCAGGTGTACACACCCAGGCCAAGGGTGGAGC
AGACCACTGTGGGATCTATGGAGATGCTTGAACGAGCAGATAACTAAGCCAAGCATGACTC
AGAGTTTCTAGAGGCCACTAGGACTGCTGAGTAATACTTGGGGGTACAGAGTCAGAAAGGA
AAGGACAAATGGTACCACTGATTAGGACCTCTGACGCTGTTTTCCCATCCTGTTTATTTGCCA
AGTGACCCTGTGCCTGTTACCTTAAGTCAAAATCTTACCCTGACAACACCTTGTACACCTGC
AGTTGGGAAGACTTTCTACATTCTCAATAAGCCACACACTTACAAAGAGAAGCAAAACCTG
GGCAGACTTGCAGCCACCAGGCCGTCTG

>Cow_MRE
CTGGCACCTATCGGGGGCCACAGCCAAGCCAGGCCTGCGCACCCGCCCAGGCAAAGGGTGG
GGCAGGCAGCTGTGAGGTCTGTGAAGATGCCCCCAGAGCTGATAACCACCTGCCGTGACTC
AGCACCCCAGGAGCCGACAGGGAGGCTGAGTCATCCCGGGTAGAGGTGGGAGGAGGGGAG
AGTGGCGGCCCCGATTAGAGCCTCCGCTTGGCCCCCCCTCCTGTTTATCTGCAGGCACGGCC
CTCGGCAAAGGTCCACAGCCAGGCTCCTGGGCCAAACCTTGACCTAGACAAAGTAAACACC
TTCATCACCTGGAGCTGGGGACACTCTAGCCCCTAGCAGATCCACATTCACAAGCAGAACTC
ACCCCCGGCAGCACACCTGTGCGGCGAGGGGCTGAG

>Opossum_MRE
CTGAAAATTATCAGTCATGTACAGACCAGGGTGAGTCATGTGCCCCAGATCAGACTGGGTG
GAGACAGCAGCTGTGAGATGTTTGTGGAGATGCTTAAAGGACCATCAAGCTCTAAAGTTAA
TAGATAAACCAGGAAACCATGACTTAGGCTTTCTGAAGGCCAAGAGTACTTGCTGAGTCATG
GCGGGGAGCTCCTGCAGAATGAAGGGGATGCAGAAGAGGGAGGATGGATCCTACTGATTGT
AACATCTAGGGCCTTCTTTGCTCCCCAACAGGTGCTCACTCCTGTTTATCTATCACAAGTAGC
CATGTACAATCCCCAAGCACAACAAAGCCTTCTAAGTCAGGATTAAATATTAATACAAAGA
ACCCAGACATTCTCTTCCTTCACTGAAAGTTAGTCTCTTATATAACATAGGGATGTCAGTGTT
TCCAAACAGCATCAACCCACCTATGAAAACAAGGAAATATC

>Platypus_MRE
GCTGCTTCCTTTCCGTTCGCCGCCTCCCGGGCTCCGCTGATGAAGCCGGGTGGGGTGACGCC
CACGGCCTGGTGTCGGGGGGTGGGCCGGGGAGCCGGGTCCGGGCCCCAGATGGCCGGGCCG
ACCCGTGACTCGGCATTTCTGGGTCTTTAGCTGAGTCACGCTCCCTGGGGGTGCCCAGAAGG
GCCCAAAAGCCAGACGGGGCCGATTGCACCATCTGTGCCCCTGCCCCGGGAGGGGCTCCTC
GTTATCTGAGGCCTGCCCGGCACCTATGGAAAACCCTGTAGCTTCAGCCTCAGCTTTTATTTA

TTTAATTTGGTAACTCAATTCCTTTCTCAGCTGTTTCACCCTGATGAACTTCCACCTCTTCCAG
AAGCAGCGTGTCCTATATTAACGTCTGTCTCGCCCCCTAGACTTCCACTAGTCAGTCAGTCA
ATTG

>Platypus_pHS#7
TACAAGATAATAAGGTTAATTTGGCCTCGCTTGAACTATGCTGTTGCAGTGAGAGGCAGTGG
GATGGCAATCAATCAATCAATGGCATTTATTGAGCATTTACTCTGGAAAGAGCACTGTACTA
AGAGTTTGGGACAGTATGCTACAATAGAGTATGTAGATATGATTCCTGTGCTCAAGGAGCTA
ACAGTGTATTGAGGAGTTTGTAGAAACAGACCTTTCTAGAAGTTCCTTCTGGCCTGAAAAAT
CTGATTCGGAGTTCTGAGCCCCTTCTAAGTCTCCAGGGTTCAAGACAGAGACAGTTGCGGTT
AAATCAGATGTGCTCCAAGCCACATGTGGAATGGCAAAATGCCGAATTGTTCTCCCCTCAGT
ATAAAAATAACCTTCAACCGCCTCTGTGGGCTACTGGGCACCAGGAGCAGACTGATGGGGC
TGGTTCAGCCCGCTCCCAACTCTGGGGATGGGAGCACAAGCTACTCTGTCCCCAGATTATCT
CCTCCTGAGTCACTGTCCCCACTCCTCAGATGCCTCTATCAGCCCAGGCCAGCCCAGGCTCA
CCCAGAGCTCCTTGGCCAAGCTATGGCCTGGGTTGGGTCCCTCCACTGTTTTTCACTGAAAA
ACAGCCAGGAGGGGGAGAAAAGGGGTGCAGGAGGAAGAAGGGGGTGGGGCAGCAGAGAGC
AACACACCAGTTCACCCCAGCCCTCTTCTTTTTTCCCCTTACCTCCCCCACCCAAGGAAAGAC
CTGTTACTCATCTGCTCTGTGACCCTGGGCAAGTCACTTAACTTCTCTGGGCCTCTTTTTCTTC
ACCACTGATAATGGTTTTCCCACCCAGTTACATTTTTTATTAGGTTCTGCAGACATCTGCAGG
GTGAGGATCCAGAGGAGAAAAGCAGAAAGTAGTGGACTGACCCAGAGGATCGGAGGGAATG
GGTAAAAGGATCATCAACTCTGGAGAGGAAAGGGTTCCAGGGAAGGTGAATAATGAAAG
AAATGAAGTGAATTTCTTC

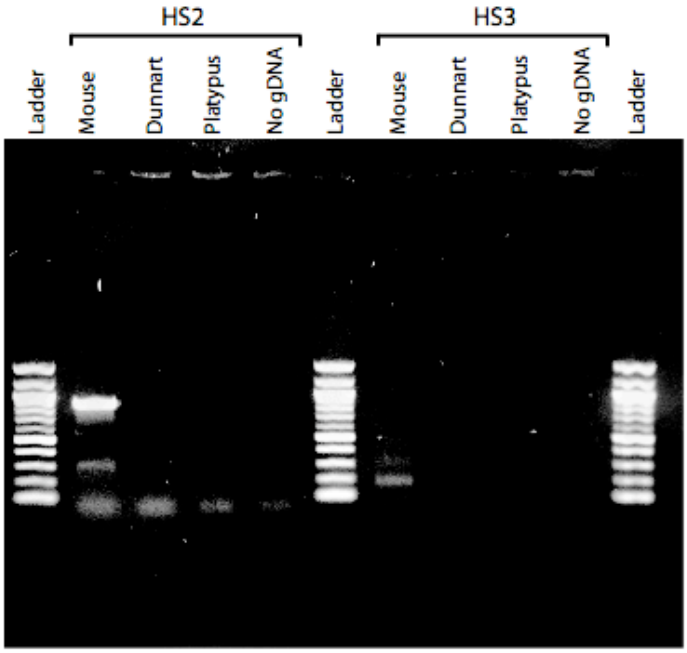
**Supplementary File 2:** The original output generated by ClustalW2 showing multi-alignment of orthologous α-globin MRE from the mouse, cow, human, opossum and platypus. In the sequence alignment, *red* boxes show two of each conserved GATA and MARE motifs.

```
CLUSTALW2


      Scores Table


SeqA Name           Len(nt)  SeqB Name           Len(nt)  Score
============================================================
1    Human_MRE      420      2    Mouse_MRE      397      72
1    Human_MRE      420      3    Cow_MRE        403      72
1    Human_MRE      420      4    Opossum_MRE    473      38
1    Human_MRE      420      5    platypus_MRE   439      23
2    Mouse_MRE      397      3    Cow_MRE        403      61
2    Mouse_MRE      397      4    Opossum_MRE    473      19
2    Mouse_MRE      397      5    platypus_MRE   439      12
3    Cow_MRE        403      4    Opossum_MRE    473      10
3    Cow_MRE        403      5    platypus_MRE   439      2
4    Opossum_MRE    473      5    platypus_MRE   439      2
============================================================
PLEASE NOTE: Some scores may be missing from the above table if the
alignment was done using multiple CPU mode. Please check the output.
```

Continued on next page

```
CLUSTAL 2.0.8 multiple sequence alignment


Mouse_MRE         -CTGGAACCTATCAGTGAC-CATAG-TCAAC---AG-CAGGTGTACACACCCAGGCCAAG 53
Cow_MRE           -CTGGCACCTATCGGGGGC-CACAG-CCAAGCCAGG-CCTGCGCACCCGCCCAGGCAAAG 56
Human_MRE         -CTGGAACCTATCAGGGAC-CACAG-TCAGCC-AGG-CAAGCACATCTGCCCAAGCCAAG 55
Opossum_MRE       -CTGAAAATTATCAGTCATGTACAGACCAGGGTGAGTCATGTGCCCCAGATCAGACT--G 57
platypus_MRE      GCTGCTTCCTTTCCGTTCGCCGCCTCCCGGGCTCCG-CTGATGAAGCCGGGTGGGGTGAC 59
                   ***     * ** *            *       * *

Mouse_MRE         GGTGG-AGCAGACCACTGTGGGA---TCTATGGAGATGCTTGAACGA------------- 96
Cow_MRE           GGTGG-GGCAGGCAGCTGTGAGG---TCTGTGAAGATGCCCCCA-GA------------- 98
Human_MRE         GGTGGAGGCATGCAGCTGTGGGGG--TCTGTGAAAACACTTGAGGGA------------- 100
Opossum_MRE       GGTGGAGACAG-CAGCTGTGAGATG-TTTGTGGAGATGCTTAAAGGACCATCAAGCTCTA 115
platypus_MRE      GCCCACGGCCTGGTGTCGGGGGGTGGGCCGGGGAGCCGGGTCCGGGCC------------ 107
                  *       *        * * *         * *          *
                                          AP1/NE-F2
Mouse_MRE         --GCAGATAACTAAGCCA----AGCATGACTCAGAGTTTCTAGAGGCCACTAGGACT-GC 149
Cow_MRE           --GCTGATAACCA--CCT----GCCGTGACTCAGCACC-CCAGGAGCCGACAGGGAG-GC 148
Human_MRE         --GCAGATAACTGGGCCA----ACCATGACTCAGTGCTTCTGGAGGCCAACAGGACT-GC 153
Opossum_MRE       AAGTTAATAGATAAACCAGGAAACCATGACTTAGGCTTTCTGAAGGCCAAGAGTACTTGC 175
platypus_MRE      --CCAGATGGCCGGGCCG---ACCCGTGACTCGGCATTTCTGG--GTCTTTAG------C 154
                        **       **       * *****  *     *     * *   **      *
                  AP1/NE-F2
Mouse_MRE         TGAGTAATACTT-GGGGGT---ACAGAGTCA--GAAA----GGAAAGGACAAATGG-TAC 198
Cow_MRE           TGAGTCATCCC----GGGT---AGAGGT-----GGGA----GGAGGGGA-GAGTGG-CGG 190
Human_MRE         TGAGTCATCCTGTGGGGGT---GGAGGTGGGACAAGG----GAAAGGGGTGAATGG-TAC 205
Opossum_MRE       TGAGTCATGGCGGGGAGCTCCTGCAGAATGAAGGGGATGCAGAAGAGGGAGGATGGATCC 235
platypus_MRE      TGAGTCACGCTCCCTGGG----GGTGCCCAGAAGGGC----CCAAAAGCCAGACGG---- 202
                  ***** *          *        *                   *   *      **
                       GATA                                                GATA
Mouse_MRE         CACTGATTAGGACCTCTGACG-----CTGTTTTCCC------------ATCCTGTTTATT 241
Cow_MRE           CCCCGATTAGAGCCTCCGC-------TTGGCCCCCC------------CTCCTGTTTATC 231
Human_MRE         TGCTGATTACAACCTCTGGTG-----CTGCCTCCCC------------CTCCTGTTTATC 248
Opossum_MRE       TACTGATTGTAACATCTAGGGCCTTCTTTGCTCCCCAACAGGTGCTCACTCCTGTTTATC 295
platypus_MRE      GGCCGATTGCACCATCTGTG------CCCCTGCCCCGGGAGGGG----CTCCTCGTTATC 252
                    * ****    * **                ***            ****  ****

Mouse_MRE         TGCCAA--GT--GAC-CCTGTGC------------CTGTTA----CCTTAAGTCAAAAT- 279
Cow_MRE           TGCAGGC-AC--GGC-CCTCGGCAAAG---GTCCACAGCCAGGC-TCCTGGGCCAAACC- 282
Human_MRE         TGAGAGGGAA--GGC-CATGCCCAAAGT--GTTCACAGCCAGGC-TTCAGGGGCAAAGC- 301
Opossum_MRE       TATCACAAGT--AGC-CATGTACAATCCCCAAGCACAACAAAGCCTTCTAAGTCAGGAT- 351
platypus_MRE      TGAGGCCTGCCCGGCACCTATGGAAAAC---CCTGTAGCTTCAGCCTCAGCTTTTATTTA 309
                  *             * * *

Mouse_MRE         CTTACCCTGACAA------CACCTTGTACACCTGC--AGTTGGGAAGACTTTCTACATTC 331
Cow_MRE           TTGACCTAGACAAAGTAAACACCTTCATCACCTGG--AGCTGGGGACACTCT----AGCC 336
Human_MRE         CTGACCCAGACAG--TAAATACGTTCTTCATCTGG--AGCTGAAGAAATTCTCTATATTC 357
Opossum_MRE       TAAATATTAATACAAAGAACCCAGACATTCTCTTC--CTTCACTGAAAGTTA----GTCT 405
platypus_MRE      TTTAATTTGGTAAC-TCAATTCCTTTCTCAGCTGTTTCACCCTGATGAACTTCCACCTCT 368
                     *       *         *         **              *

Mouse_MRE         TCAATAAGCCACACACTTACAAAGAGAAGCAAAACCTGG--GCAGACTTGCAGCCACCAG 389
Cow_MRE           CCTAGCAGATCCACATTCACAAGCAGAACTCACCCCCGGCAGCACACCTGT-GCGGCGAG 395
Human_MRE         TCATCAGATCCCATGTTCACAAACAGAACTCACACCTGG--ATGGACTTACAACAAGGCA 415
Opossum_MRE       CTTATATAACATAGGGATGTCAGTGTTTCCAAACAGCATCAACCCACCTATGAAAACAAG 465
platypus_MRE      TCCAGAAGCAGCGTGTCCTATATTAACGTCTGTCTCGCCCCCTAGACTTCCACTAGTCAG 428
                                  *                     ** *

Mouse_MRE         GCCGTCTG--- 397
Cow_MRE           GGGCTGAG--- 403
Human_MRE         GCATC------ 420
Opossum_MRE       GAAATATC--- 473
platypus_MRE      TCAGTCAATTG 439

PLEASE NOTE: Showing colors on large alignments is slow.


      Guide Tree


(Human_MRE:0.03981,(Mouse_MRE:0.17080,Cow_MRE:0.21459):0.04375,(Opossu
m_MRE:0.43585,platypus_MRE:0.53909):0.16200);
```

**Supplementary File 3:** PCR amplification of HS2 and HS3 in mouse, fat-tailed dunnart and platypus gDNA by replicating exact materials and methods by De Leo et al.(2005). The size of bands was calibrated with 100 bp DNA ladder (New England Biolabs Inc). Mouse gDNA was used as positive control while no gDNA was used as a negative control. For HS2 amplification, two bands of approximate sizes 300 bp (faint) and 900 bp (bright) were seen for mouse only, while for HS3 amplification, only one faint band of ~200 bp was seen for mouse. Sequencing confirmed the presence of mouse HS2 at ~300 bp and HS3 at ~200 bp, while the bright 900 bp seen in well containing mouse HS2 was due to non-specific amplification from mouse chromosome 11. Note: No bands were observed from the platypus or dunnart gDNA.



# Chapter 6: Synopsis and Future Work

The three research chapters that form the core of my thesis, along with the results from Appendix 1, serve a common goal; to understand how the complex α- and β-globin genes, their functions, clusters and regulatory regions evolved in amniotes. My findings have had a major impact on our understanding of globin gene evolution. In this final chapter, I discuss the significance of my findings with respect to other published studies, discuss their implications and suggest future directions.

## Timing and order of globin gene duplications

After the duplication of a single primordial globin gene into ancestral α- and β-globin genes in the same genomic region, which occurred around 450-500 MYA in the ancestor of jawed vertebrates, there were independent repeated rounds of duplications and losses of these globin genes in the teleost and amphibian lineages that expanded their α-β globin clusters (Goodman et al., 1975). In amniotes, however, the α- and β-globin genes separated onto two different chromosomal locations, raising important questions about their evolution.

My study of the α- and β-globin clusters of the platypus (Chapter 2) and two lizard species (Chapter 4) have not only filled in missing pieces of information about their gene content and organization, but also clarified the timing and order of globin gene duplications and losses in amniotes, in particular, before and after monotreme/therian divergence. These data led to the formulation of a new model for globin gene evolution. My studies of expression of α- and β-globin genes also provided insight into the evolution of function.

### Alpha-globin gene evolution

The bird α-globin clusters contain three genes arranged in the temporal order in which they are transcribed *HBP-HBK-HBA* (Engel and Dodgson, 1980, Dodgson et al., 1981, Alev et al., 2009). The *HBP* is expressed in embryos, *HBK* in both embryo and adult, and *HBA* in adults only. Prior to this study, the arrangement and order of genes in the α-globin clusters of non-avian reptiles was unknown, so no comparisons of birds and

reptiles were available to assist in the construction of ancestral globin clusters in their common amniote ancestor. Therefore I attempted to characterise the globin cluster in the green anole lizard (*Anolis carolinenesis*) using the draft genome assembly (anoCar1 Feb. 2007).

My results in Chapter 4A showed that the green anole contained *HBK* and *HBA*, with a confirmed expression of *HBK* in dewlap and brain, organs not expected to undergo globin synthesis. However, the two genes were located on different scaffolds flanked by different markers; *HBA* is located on a short scaffold of ~15 kb, whereas *HBK* is located on a scaffold of ~167 kb, which contains a large number of gaps, raising questions in regards to the assembly of the region containing these genes. Although the separation of *HBK* and *HBA* could be due to the incomplete assembly of the green anole genome, if the separation is real, it would imply a unique rearrangement between the two α-globin genes. Embryonic *HBP* could be found nowhere in the genome sequences. Since *HBP* is present in the amphibian outgroup, the three-gene (*HBP-HBK-HBA*) cluster must have existed in the common ancestor of reptiles and mammals, but *HBP* was secondarily lost in the lizard. These results are consistent with those reported by Hoffmann et al. (2010b).

The rearrangement that separated *HBK* and *HBA* in the green anole does not seem to be general for lizards, as I showed in Chapter 4B that the Australian bearded dragon lizard (*Pogona vitticeps*) *HBK* and *HBA* are located together on the same BAC clone, flanked by the genes *C16orf35* and *GBY*. These same genes flank the α-globin locus in mammals, as well as the α-β globin locus in the frog *Xenopus tropicalis*. Until the completion of the green anole genome assembly, or sequencing of other non-avian reptilians, the α-globin gene content and order will remain uncertain in these groups.

My study of the platypus globin clusters (Chapter 2) clarified how the increasing complexity of the globin loci in mammals evolved. I demonstrated the presence of six transcriptionally active α-like globin genes in the platypus, arranged in the order *HBZ-T1, HBZ-T2, HBK, HBA-T1, HBA-T2, HBA-T3*, which is similar to the arrangement in marsupials and eutherians. However, the platypus genome lacks *HBQ* (the θ-globin gene). Due to the inconsistencies in nomenclatures, two mammalian *HBZ* are orthologous to reptilian embryonic *HBP*, and three mammalian *HBA* genes are orthologous to reptilian adult *HBA*.

I studied the expression of all six globin genes in adult platypus tissues, but was unable to obtain tissues from different stages of platypus development (especially the embryo). My expression data from adult platypus tissues showed that all six α-like globin genes were expressed in adults. Some were expressed in erythroid cells only (*HBK* in spleen), while others were expressed in both erythroid and non-erythroid cells (*HBZ-T1,2* in spleen and testis; *HBA-T1,2,3* in spleen, testis, liver, kidney, lung and brain). The expression of these genes in spleen makes sense because this tissue is a major site of erythropoiesis in platypus (Tanaka et al., 1988). However, the expression of these genes in non-erythroid cells was surprising and interesting because expression is limited to erythroid tissues in other vertebrate groups (amphibians, birds and therian mammals). This suggests that some globin genes acquired a broad expression in the monotreme lineage, perhaps in response to the platypus' aquatic and terrestrial lifestyle and the hypoxic conditions of a confined burrow. It also suggested that these genes may have functions outside of binding and transporting oxygen.

Comparisons of the α-globin clusters in reptiles, platypus and other mammals clarified the timing and order of duplications of α-like globin genes in amniotes. After the divergence of reptiles and mammals (315 MYA), but before the divergence of monotremes and therian mammals (166 MYA), the *HBP*-globin gene duplicated once and the *HBA*-globin gene duplicated twice to result in a six-gene (*HBZ-T1, HBZ-T2, HBK, HBA-T1, HBA-T2, HBA-T3*) cluster in the ancestral mammal, in which all genes were transcriptionally active and functional (Figure 1). However, during the course of evolution, the two *HBZ* and three adult *HBA* were homogenised by ongoing gene conversion events, leading to the gene tree that does not match the duplication history of the individual genes. The close similarity of the platypus *HBA-T1* and *HBA-T3* sequences suggests that a very recent gene conversion event homogenised their sequences.

After the divergence of monotremes from therian mammals, *HBA-T2*-globin gene duplicated to produce *HBQ*, resulting in a seven-gene (*HBZ-T1, HBZ-T2, HBK, HBA-T1, HBA-T2, HBA-T3, HBQ*) cluster (Figure 1). This arrangement is seen in contemporary therian mammals, except for where there are some gene losses and duplications, and psuedogenisations of some of these genes.

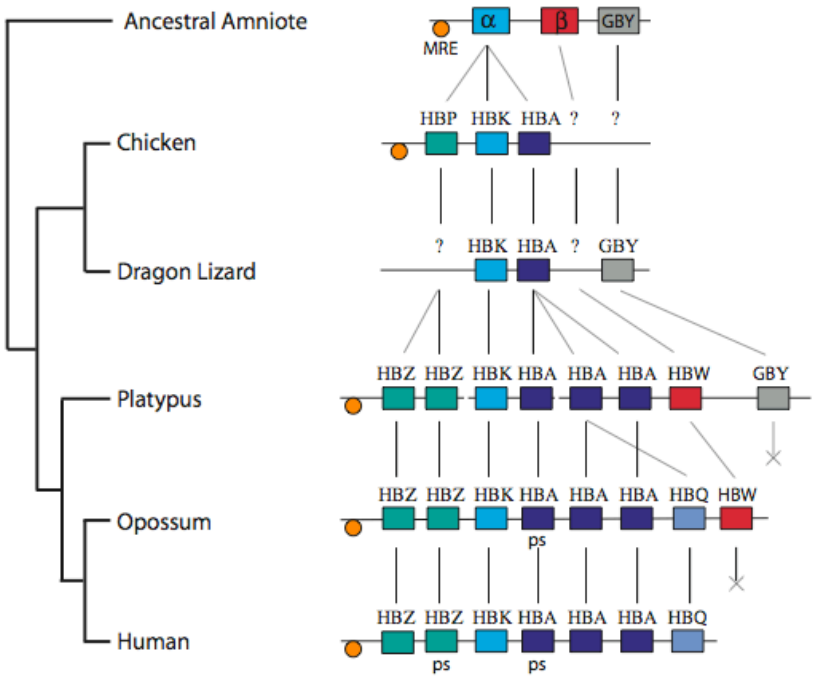



**Figure 1: Evolution of α-globin genes in amniotes**

In the amniote ancestor, the α-globin gene lay next to β and *GBY* globin genes. During the evolution of reptiles and mammals, the ancestral α-globin gene underwent repeated rounds of duplications to maintain its function, but there were no further duplications of β or *GBY*, so these genes are absent from some lineages. The coloured boxes show orthologies among genes. Note: The figure is not drawn to scale.

## Identification of two additional globin genes in the ancestral mammalian α-globin cluster

Characterisation of the platypus globin genes also revealed the presence of a β-like globin gene called ω-globin (*HBW*), and another globin gene family member called Globin Y (*GBY*) located 3′ to the α-globin cluster. Orthologues of *HBW* and *GBY* have been previously discovered, respectively, 3′ to the α-globin cluster in marsupials (Wheeler et al., 2001) and frogs (Fuchs et al., 2006).

Identification of the orthologues of these two genes together in the platypus α-globin cluster is important because their presence supports the hypothesis that the α-β-*GBY* arrangement was in place in the common ancestor of mammals and frogs, and has been

conserved for more than 354 million years. Some genes must, therefore, have been secondarily lost in some tetrapod lineages. For example, , I could find no orthologues of *HBW* in either birds or lizards, suggesting it was lost early in the reptilian lineage. Although I found *GBY* present in lizards at the expected location, it was not found in chicken, implying that *GBY* was secondarily lost in the bird lineage. Likewise, in mammals, *GBY* was lost in therians before their divergence from monotremes, and *HBW* was lost in eutherians after their divergence from marsupials.

Both *HBW* and *GBY* genes are transcriptionally active. The expression pattern of *GBY* is similar in frogs (Fuchs et al., 2006) and platypus, being broadly expressed in almost all tissues, but most strongly in the gonad of both species. This suggests that a special function in respiration or regulation exists in the gonads, although there is no direct evidence for such a function.

In contrast, *HBW* has quite different expression profiles in the marsupial and platypus. It is expressed just before and after the birth of a marsupial newborn (Holland and Gooley, 1997, Holland et al., 1998). In platypus, although tissue samples from early stages of development were unavailable, *HBW* is expressed in the adult spleen. The *HBW* is known to bind with α-globin proteins to form a functional haemoglobin molecule to transport oxygen in marsupial neonates (Holland and Gooley, 1997, Holland et al., 1998), but its function in platypus spleen is not known.

**Beta-globin gene evolution**

My study into the platypus and lizards β-globin genes was very useful for reconstructing the β-globin gene model in amniotes that evolved from a single ancestral β-globin gene and duplicated independently in the reptilian and mammalian lineages (Hoffmann et al., 2010b). I found two β-like globin genes in the green anole genome, which was consistent with the number present in snakes and tuatara, but not with the single gene in crocodiles and turtles or the four in birds. Phylogenetic analyses presented in Chapter 4A grouped both lizard β-like globin genes together with other lizard *HBB1* and *HBB2* genes, suggesting a lizard-specific duplication of a β-globin gene. This was also the conclusion of an independent study on the green anole lizard β-globin genes performed by Hoffmann et al. (2010b).

My phylogenetic analysis also showed species-specific clustering of the β-like globin genes, in which β-like globin genes of lepidosaur (tuatara, snakes and lizards), birds and crocodiles each formed monophylic clades. The lepidosaur β-globin genes were divided into two main clades; one contained *HBB1* of snakes and tuatara, and the other was a sister clade that contained *HBB1* and *HBB2* of lizards as well as *HBB2* of snakes and tuatara. These results led me to propose two alternative models for the evolution of β-globin genes in reptiles. Since crocodiles and turtles contain one β-globin gene, one model proposed that the reptilian ancestor contained a single ancestral β-globin gene, which first duplicated in the lepidosaur lineage to result in proto-$\beta^1$ and proto-$\beta^2$ globin genes, whose descendants are both seen in tuatara and snakes. During the evolution of lizards, however, the proto-$\beta^1$ gene was secondarily lost from their genome, while proto-$\beta^2$ gene underwent tandem duplication to result in two paralogous β-globin genes in all lizards, as supported by my phylogenetic analysis. In the bird lineage, there were further duplications of the ancestral β-globin gene (independent from the duplication in lepidosaurs) to result in a four-gene (*HBR, HBB-T1, HBB-T2, HBE)* cluster in chicken and zebrafinch.

The other possible, but less parsimonious, model proposed that the reptilian ancestor possessed two proto-β-globin genes, which underwent differential gene loss and duplication in different reptilian lineages. The proto-$\beta^1$ is found in crocodiles, turtles, tuatara, and snakes, and was duplicated further in birds, whereas in lizards it was secondarily lost from their genomes (Figure 2- Model B). The proto-$\beta^2$ is found in tuatara and snakes, and as two paralogous copies in lizards, but was secondarily lost from the common ancestor of birds, crocodiles and turtles.

However, data from other reptilian species (e.g. tuatara, crocodiles, turtles, snakes) are lacking. The nucleotide sequences of all globin genes and their expression profiles would provide a clearer view of globin gene evolution in reptiles. Thus, in future, it would be important to characterise the entire α- and β-globin clusters in more reptiles to gain further insight into the gene content, organization, function and evolution of amniote globin genes.

Like the marsupial β-globin cluster, the platypus β-globin cluster contained two β-like globin genes, in the order 5′-*HBE-HBB*-3′, located on platypus chromosome 2q5.1.

Both genes are transcriptionally active, with *HBE* expressed in the adult spleen and *HBB* (like *HBA-T1, 2,3*) expressed in all adult tissues.

Phylogenetic analyses presented in Chapter 2 showed that the platypus *HBE* was more closely related to the platypus and echidna adult *HBB* than to therian *HBE*. However, the location, amino acid and expression profile of *HBE* support its orthology with the marsupial and eutherian embryonic *HBE*. This suggests that the ancestral β-globin gene duplicated into proto-*HBE* (early-expressed) and proto-*HBB* (late-expressed) globin genes just prior to monotreme and therian divergence (>166 MYA) but the sequence of the platypus *HBE* may have been homogenised by gene conversion events, leading it to group with other monotreme adult *HBB* (Figure 2). In marsupials, the progenitors of both proto-*HBE* and proto-*HBB* are present in a transcriptionally active state, whereas in eutherians these genes duplicated further to form the more complex clusters seen in contemporary eutherian species (Figure 2).

However, Opazo et al. (2008b) reported that the platypus *HBE* and *HBB* were not 1:1 orthologues of therian *HBE* and *HBB*, respectively. They suggested that the platypus β-like globin genes arose by independent duplication of an ancestral β-globin gene in the monotreme lineage, with a separate duplication event, just prior to the divergence of marsupials and eutherians, producing the progenitors of *HBE* and *HBB* of therians (Figure 2). Thus, the β-globin genes of reptiles, monotremes and therians are not orthologous, having arisen by lineage-specific independent duplications of ancestral β-globin gene (Hoffmann et al., 2010b). However, it is difficult to see how three independent duplications could lead to β-globin genes having similar expression in erythroid-specific cells and different stages of development.

Support for this hypothesis of independent duplication of an ancestral β-globin gene in monotremes is provided by my study of regulatory regions in Chapter 5 that showed that platypus sequences contain no orthologues of the bird or therian *cis*-regulatory regions. This suggests that the regulatory regions also evolved independently along with the β-globin genes they regulate in the monotreme, therian and reptilian lineages. The developmental stage-specific erythroid expression of β-like globin genes in birds and therians can be explained by either convergent evolution, that resulted in the performance of similar functions by independently derived regulatory elements, or alternatively by major regulatory support coming from other regions, such as α-globin

MRE, that would coordinate the timing and expression of both the α- and β-globin genes in an erythroid- and developmental stage-specific manner.

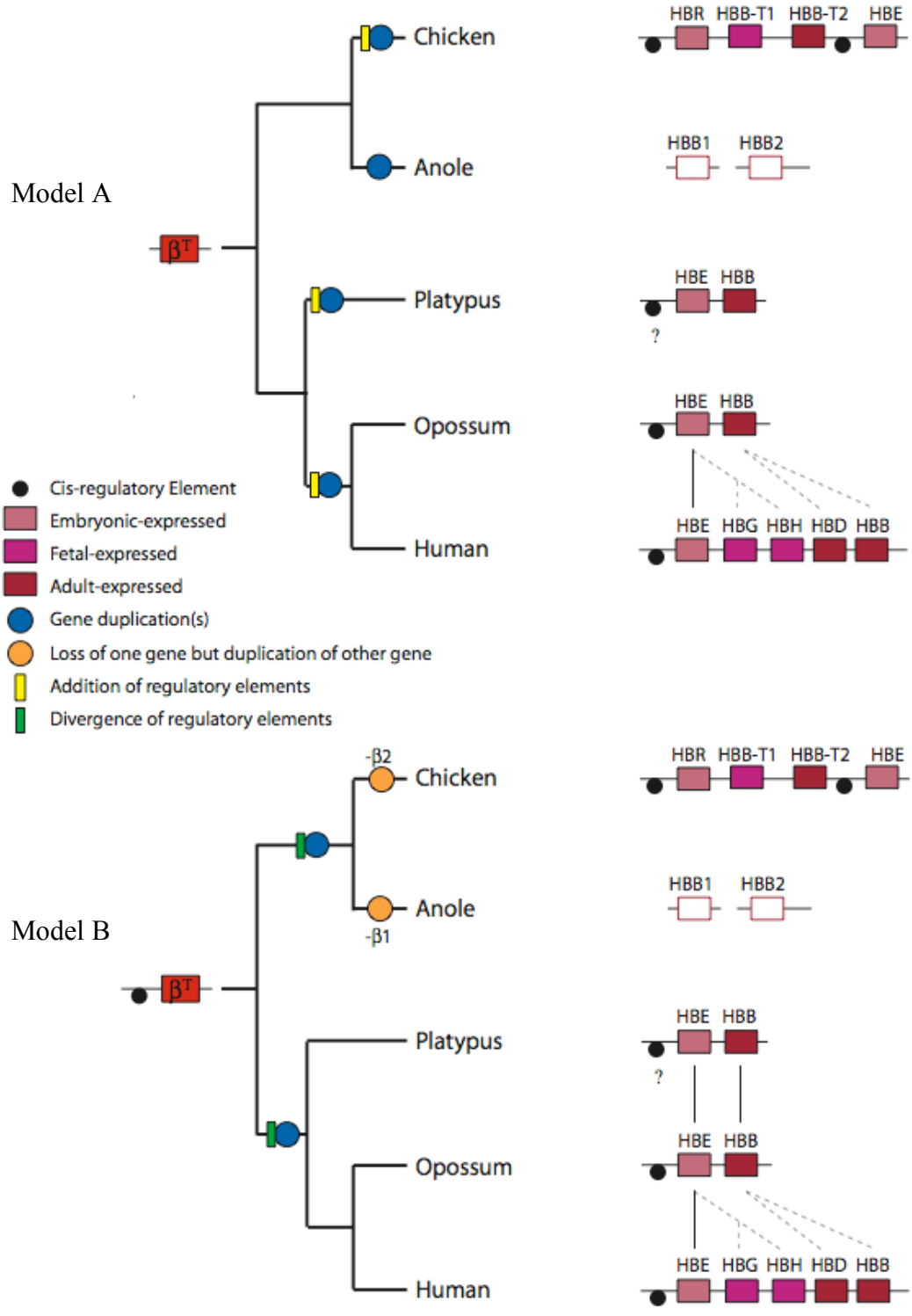


**Figure 2 (previous page): Alternative models for the evolution of β-globin genes in amniotes**

According to Model A, the ancestral transposed $\beta^T$-globin gene duplicated independently four times in the amniote lineage. It may have not brought any regulatory elements and as a consequence, regulatory elements also evolved independently. According to Model B, the ancestral transposed $\beta^T$-globin gene duplicated twice in amniotes, one in the ancestor of reptiles and the other in the ancestor of mammals, followed by further duplications, gene losses and conversions. It also predicts that some regulatory regions were carried with the transposed genes, which underwent further divergence and duplications in reptiles and mammals. Even though Model B is more parsimonious than Model A, phylogenetic data suggest many independent duplications of ancestral β-globin gene in amniotes, favouring Model A. Note: The different colours of these genes do not imply orthology unless specified by dashed lines between genes. The lizard globin genes are not colour-coded as their expressions have not yet been studied. The figure is not drawn to scale.

## The globin clusters are a part of large evolutionarily conserved regions

In chapter 2 I described comparative studies of the globin flanking regions, in the platypus and other jawed vertebrates, which suggests that the α- and β-globin loci are each part, not only of a small conserved region, but also of large evolutionarily conserved blocks.

Using the fine resolution of the platypus α-globin BAC clone sequence, I was able to demonstrate that *MPG*, *C16orf35*, *GBY* and *LUC7L* flank the platypus α-β globin locus in a ~150 kb region. This region was also found to be present in the genomes of human (spanning ~155 kb), opossum (~250 kb), chicken (~165 kb), frog (>250 kb), zebrafish (~30 kb) and pufferfish (~40 kb). All of these regions contained the same set of genes arranged in the same order *MPG-C16orf35-(α-β)-GBY-LUC7L,* although some genes were secondarily lost in some lineages. Likewise I found that in the genomes of lizards (Chapter 4), α-globin clusters are flanked by orthologues of the same genes (*MPG*, *C16orf35*, *GBY* and *LUC7L*). This helped me to reconstruct the arrangement in the common ancestor of reptiles in the order *MPG-C16orf35-(α)-GBY-LUC7L*, with *GBY* lost in birds and the β-like *HBW* lost in all reptiles.

In turn, this conserved segment is part of a much larger region flanked by *POLR3K* and *AXIN1,* synteny of which is conserved in 22 species that diverged 500 million years ago. This ancestral region (extending from ~135 kb to ~248 kb in the human genome) contained all the α-globin genes, their *cis*-acting elements and flanking genes (Flint et al., 2001, Hughes et al., 2005). This implies that the genomic context of the amniote α-globin clusters are the same as fish and frog α-β globin clusters, and that their location in a much larger evolutionarily conserved region implies that the whole region is ancient, and has been preserved over a long evolutionary time (Figure 3).

In contrast, the flanking region of the amniote β-globin cluster points towards quite a different origin. I found that the platypus β-globin cluster, like the β-globin clusters of other mammals and birds, shared no flanking genes with the platypus α-globin cluster, or with the frog and fish α-β globin clusters (Bulger et al., 1999, Gillemans et al., 2003, Hardison, 2005), as would be predicted by a hypothesis that this cluster originated by whole genome or regional duplication of the α globin cluster. Instead, all amniote β-globin clusters, spanning 315 million years of evolution, were embedded in a large region of olfactory receptor (*OR*) genes flanked by singleton genes *ILK, CCKBR* at the 5′ end and *RRM1* at the 3′ end. This ancient arrangement has been broken up in some groups; for instance in teleost fish and frogs, I found *ILK*, *CCKBR* and *RRM1* to map separately to distinct chromosomes, distant from any α or β globin genes.

The genomic context of the amniote β-globin clusters within ORG regions and distant from α-globin clusters suggests insertion of an ancestral globin gene within a different large evolutionary conserved region containing *ILK-CCKBR-OR-(β)-OR-RRM1* in the common ancestor of amniotes (Figure 3). This order has been conserved in almost all contemporary amniotes, except for dragon lizard, in which I found a breakpoint between *CCKBR* and the β-globin cluster that interrupted this region. Without sequence from other lizards, it is not possible to determine at what point this rearrangement occurred, and without a refined search, it is not possible to determine the outer limits of this conserved region.

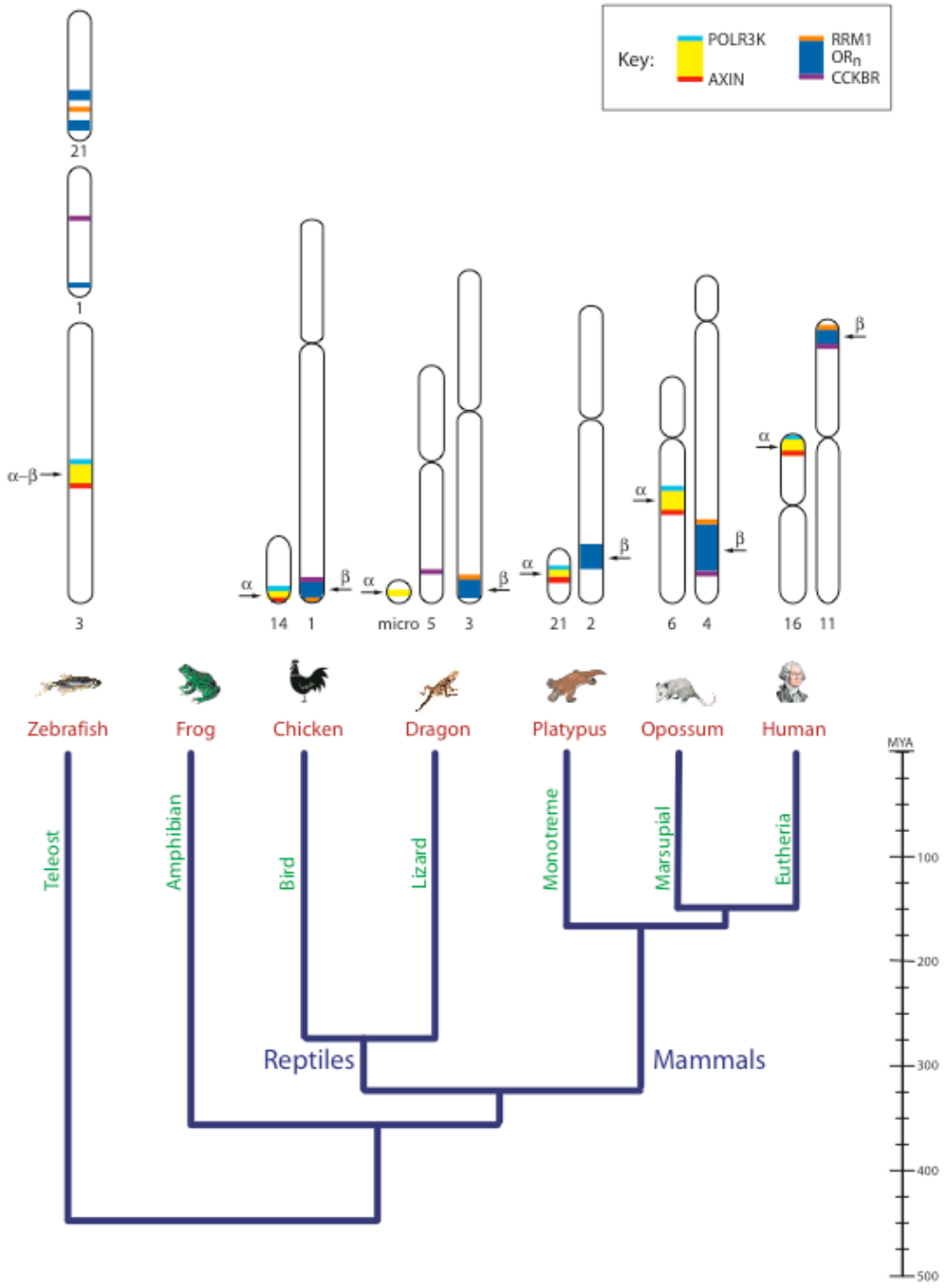



**Figure 3: Conserved synteny across jawed vertebrates**

Identification of the conserved region of synteny is based on flanking genes. The region shown in *yellow* between flanking genes *POLR3K* (*light blue* box) and *AXIN1* (*red*) contains the ancestral region *MPG-C16orf35-*$\alpha_n$*-*$\beta$*-GBY-LUC7L*. Some genes in this region were deleted, duplicated or rearranged in some lineages (e.g. zebrafish). The region (*blue*) between flanking genes *CCKBR* (*purple*) and *RRM1* (*orange*) contains $OR_n$-$\beta_n$-$OR_n$ in amniotes, but not in fish. Mapping data from frog is lacking. Note: Chromosomes and conserved regions are not drawn to scale.

**The *ILK-CCKBR-OR-(β)-OR-RRM1* region is conserved in marsupials but has received another transposition**

At the time this study commenced, the genome of only one marsupial species, the Brazilian short-tailed opossum (*Monodelphis domestica*) was being sequenced, but its physical map was very incomplete. The opossum genome browser (monDom4 Jan. 2006 in the UCSC Genome Browser) had not assigned the location of β-globin cluster to any of the chromosomes, although *RRM1* was assigned to chromosome 4 and *CCKBR* to chromosomes 4 and 5. However, I found that, unexpectedly, in the site between *CCKBR* and *RRM1* lay an orthologue of another gene *FTSJ1* that is located on the X chromosome in humans. From this analysis, it is uncertain whether the *ILK-CCKBR-OR-(β)-OR-RRM1* arrangement is conserved in any marsupial species.

To determine whether this arrangement was conserved in marsupials, I investigated the genome of an Australian model marsupial, the tammar wallaby (*Macropus eugenii*). The tammar genome was next in the line for sequencing, but at a much lower depth of coverage (2x) compared to opossum (6x). Finding the locations of the β-globin cluster and its flanking genes was not only important to this study, but would contribute to building the physical map for the tammar wallaby sequencing project. To investigate globin gene arrangement and location in the tammar, I first screened the tammar BAC library for β-globin and the flanking genes *ILK, CCKBR, RRM1* and *FTSJ1*, and then mapped the positive BAC clones containing these genes onto the metaphases using FISH. The results of this study are presented in a published manuscript by Deakin et al. (2008) (Appendix 1).

My FISH results showed that the β-globin cluster, *ILK*, *CCKBR*, *RRM1* and *FTSJ1*, all mapped to the long arm of tammar chromosome 5 at 5q3 (Figure 4), suggesting that the β-globin cluster was indeed flanked by *ILK*, *CCKBR* and *RRM1* in this marsupial species. This is consistent with the hypothesis that insertion of the β-globin gene into an *ORG*-containing region occurred earlier and was retained in marsupials.

The unexpected presence of *FTSJ1* immediately adjacent to the β-globin cluster in the tammar suggested a recent insertion of this gene into the *ILK-CCKBR-OR-(β)-OR-RRM1* region early in the marsupial lineage. This was confirmed when I later observed the same set of genes in the order *ILK-CCKBR-OR-FTSJ1-OR-(β)-OR-RRM1* in the

updated genome browser of opossum (monDom5 Oct. 2006). Therefore, results from both the tammar and opossum suggested that the insertion of *FTSJ1* between *CCKBR* and β-globin cluster occurred at least 55-80 MYA, since the American and Australian marsupial lineages both show this order.

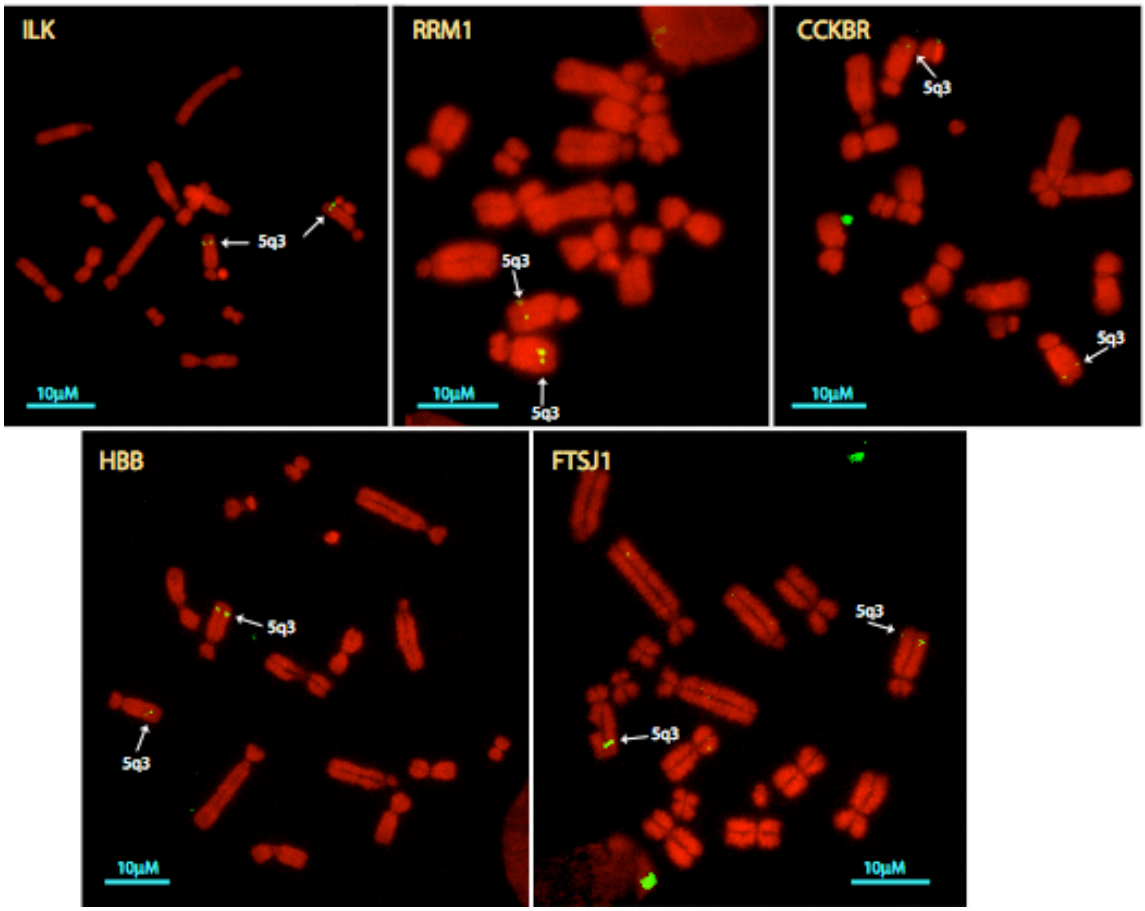



**Figure 4: Chromosomal location of *ILK*, *RRM1*, *CCKBR*, *HBB* (β-globin) and *FTSJ1* in the tammar wallaby.**

Single colour fluorescence in situ hybridisation showing locations of these genes onto 5q3 (green). Chromosomes are counterstained with DAPI (red).

### Contradictory localizations of tammar β-globin resolved

The high resolution and efficiency of the mapping techniques used here not only provided the first accurate localization of β-globin genes in the tammar, but also cleared up confusion from contradictory mapping data in the literature. This resolution is important because the correct map positions provide insight into chromosome homology across other marsupial species.

My FISH mapping of β-globin-containing BACs in the tammar completely contradicted the early mapping data of Sinclair and Graves (1991). Sinclair used a β-globin (*HBB*) cDNA from a dasyurid marsupial, *Didelphis viverrinus,* as a probe to map the location on tammar chromosomes using radioactive *in situ* hybridisation (RISH). These authors also mapped two other markers, *HRAS* (cellular oncogene homolog of Harvey rat sarcoma virus V-Haras 1) and *CAT* (catalase), which are syntenic to *HBB* on human chromosome 11p. The *HRAS-HBB-CAT* genes appear to be part of a conserved synteny group found in most eutherian mammals. Sinclair reported that all three genes were located in a region on the long arm of chromosome 3 (3q3.1) in the tammar wallaby (Figure 5A), suggesting that *HRAS-HBB-CAT* synteny is also conserved in marsupials.

However, these early results were inconsistent with the more recent, but indirect study by De Leo et al. (2005). They cloned a BAC containing the β-globin cluster from the fat-tailed dunnart (*Sminthopsis crassicaudata*) and mapped it to dunnart chromosome 3q. Although De Leo and colleagues did not clone or map the tammar β-globin gene directly, they were able to predict its location from her study of chromosome homology between marsupial species. They used DNA "chromosome paints" from flow sorted tammar chromosomes to detect homologies in dunnart chromosomes, discovering that dunnart chromosome 3q is homologous to tammar chromosome 5 (De Leo et al., 1999). Hence, De Leo et al. (2005) predicted the location of the β-globin cluster to be on tammar 5q rather than 3q (Figure 5B).

The present FISH mapping results (Appendix 1) confirmed directly that the β-globin cluster indeed lies on tammar 5q3 (Figure 5C). This confirmed the prediction by De Leo et al. (2005) and refuted the conclusions of Sinclair and Graves (1991). Thus, chromosome homology between closely related species allows reliable and consistent predictions of gene position. This is an important point to establish, because full genome sequences are now available for only two model marsupial species, the tammar wallaby and the opossum. However, using the painting homology right across all marsupial groups established by Rens et al. (2004), it should be possible to predict the location of any gene in any marsupial species.

My results also implied that RISH mapping using small heterologous probes provides unreliable and inconsistent gene locations. There are several factors that make RISH

less sensitive and efficient. Firstly, Sinclair and Graves (1991) used cDNA probes from distantly related species (*D.viverrinus* and human). Since the level of sequence conservation decreases as the phylogenetic distance increases, heterologous probes hybridise much more weakly to the target gene region, reducing the binding between probe and target site, and decreasing the sensitivity of hybridisation. The cDNA probes, as well as being heterologous, were much smaller than the BAC clones used here, so the amount of probe DNA bound by the target is much smaller and the signal is weaker, so that detection is much less sensitive. The lower efficiency of radioactive decay, means that detection must rely on counting radioactive grains over all chromosomes in many cells (typically more than 100), thus background accumulates and statistical analysis must be used to demonstrate that the signal peak is significantly above background (Ewens et al., 1992). These factors undoubtedly compromised the sensitivity and efficiency of the older physical gene mapping by RISH. This is important because most of the older gene assignments relied on RISH data, and there has been a serious question about whether they should be included in a comparative map of this important model species (Alsop et al., 2005).

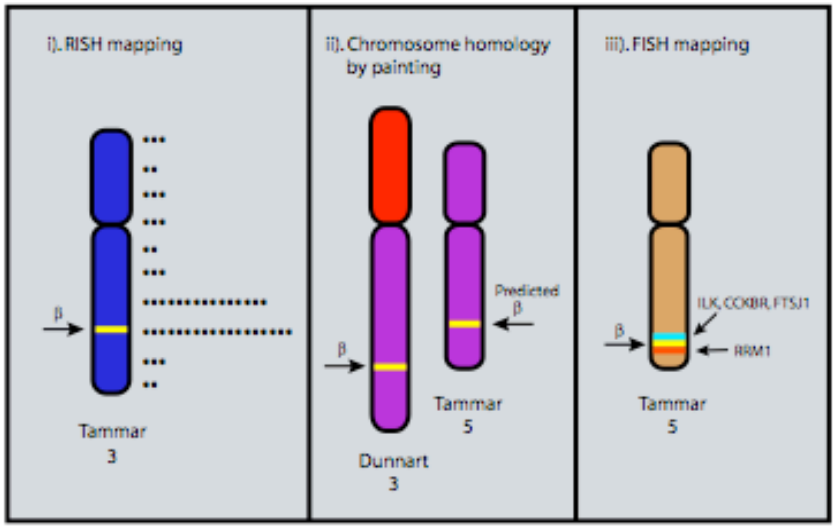



**Figure 5: The β-globin localization in the tammar wallaby using different mapping approaches.**

i) Localization of β-globin by hybridising a radioactive *D.viverrinus HBB* cDNA probe to the tammar 3q using RISH ii) Prediction of β-globin location by chromosome painting, showing that dunnart 3q (where dunnart β-globin maps) is equivalent to tammar 5. iii) In the present study, the location of the β-globin cluster was directly determined by FISH to be near the terminus on tammar 5q3, near *ILK, CCKBR, RRM1* and *FTSJ1*.

## Evolution of α- and β-globin clusters in amniotes

The study of globin gene clusters in the platypus, green anole, dragon lizard and tammar wallaby allowed me to achieve the main aim of this project, that is, to challenge the hypotheses for the evolution of α- and β-globin clusters in amniotes.

Jefferys et al. (1980) were the first to put forward a theory based on the observation that amphibians, as an outgroup, possessed a cluster containing both α- and β-like globin genes, but amniotes possessed distinct clusters for these globin genes. According to their hypothesis, ancestral α and β globin genes separated before the evolution of amniotes as the result of a break between both genes (chromosome fission), perhaps followed by translocation. This theory had been widely accepted for more than two decades until research into marsupials suggested a radically different model for globin gene evolution.

The theory that had gained general acceptance by the time my work was initiated was put forward by Wheeler et al. (2001 and 2004) on the basis of their finding of a β-like *HBW* adjacent to the α-globin cluster. The authors suggested that the amniote α- and β-globin clusters evolved from a duplication (segmental, chromosomal or whole genome) of a region that contained the ancestral α-β region. The two copies of the region underwent differential gene loss in bird and mammalian lineages. According to this model, the bird β-globin clusters were paralogous to the mammalian β-like globin clusters.

However, my observations of the phylogenetic and flanking analyses contradict this model because markers that flank α- and β-globin clusters are completely different in all amniotes.

The data presented in Chapter 2 clearly showed that:

1) The genome of the platypus, like those of marsupials, contains a β-like *HBW* adjacent to the α-globin cluster, which is not present in any reptile or eutherian mammal.
2) Sequences of marsupial and monotreme *HBW* group together as an outgroup to all bird and mammalian β-like globin genes, suggesting that *HBW* is more ancestral than amniote β-like globin genes.

3) The bird β-like globin genes grouped as a sister to mammalian β-like globin genes, suggesting an orthologous, rather than paralogous, relationship of the bird and mammalian β-globin clusters. This conclusion is consistent with the results of Aquileta et al. (2006a) and Opazo et al. (2008b) whose work also produced similar phylogenetic trees but lacking the platypus data.
4) The flanking regions of the amniote α- and β-globin clusters are totally different; α-globin resides in an ancestral region that has been conserved for about 500 million years, whereas β-globin was inserted quite recently into a region containing *OR* genes that has been conserved for some 315 million years. This, too, provided evidence that the bird β-globin cluster was orthologous (not paralogous) to the mammalian β-globin cluster.

These results clearly refuted the *in trans* duplication theory by Wheeler et al. (2001 and 2004). Rather, my results suggested an alternative model for the evolution of α- and β-globin clusters in amniotes, which is much simpler than previous models.

**Transpositional model for the evolution of the amniote β-globin cluster**

According to the transpositional model, all amniote α-globin clusters were derived from the ancestral region containing *MPG-C16orf35-(α-β)-GBY-LUC7L*, that has been conserved since the evolution of jawed vertebrates (410 MYA), except for their loss in some species of genes *HBW* and/or *GBY*. However, the amniote β-globin cluster originated in an ancestral amniote just before divergence of reptiles and mammals (315 MYA) by transposition of a single copy of the β-globin gene from this ancient region, and was inserted into another region containing a sea of *OR* genes flanked by *ILK*, *CCKBR* and *RRM1*. It then duplicated and diverged further to result in β-globin clusters in contemporary reptiles, birds and mammals.

This model was further elaborated by studies of the genome of the dragon lizard (Chapter 4B) that showed the presence of two unlinked clusters; the *C16orf35-(α)-GBY* on a pair of micro-chromosomes, and the *(β)-OR-RRM1* on 3q. In Chapter 3, I outlined some possible ways in which this transposition could have taken place.

A particularly interesting question raised by the transpositional hypothesis is the effect that transposition might have had on genes and regulatory regions.

**Effect of transposition on regulatory regions**

The transposition model easily accounts for the evolution of α- and β-globin genes in amniotes, but questions surrounding the evolutionary origins of the regulatory regions for the α- and β-globin loci in amniotes are difficult to answer.

In Chapter 5, I reported investigations of this question further by characterising the *cis*-regulatory regions of the platypus α- and β-globin clusters. I found the platypus α-globin major regulatory element (MRE) in the fifth intron of an adjacent gene, *C16orf35*, conserved with those of other mammals, chickens and teleost fishes. This suggests that the regulatory region evolved at the same time as α- and β-globin genes first duplicated from an ancestral primordial globin gene about 500 MYA (Flint et al., 2001). The MRE has since been conserved to some extent in sequence, transcription factor binding sites and its ability to enhance gene expression amongst jawed vertebrates, reflecting the ancient origin of the regulatory, as well as the structural genes, in the α-globin locus.

Conversely, exhaustive searches for a platypus β-globin regulatory region failed to identify a regulatory element conserved either between platypus and therian mammals, or platypus and chicken. Instead, I found a unique hypersensitive site that I suggest is a part of the regulatory region for platypus β-like globin genes The available data do not allow me to distinguish whether the β globin gene carried some regulatory sequences when it transposed, which then duplicated and diverged, or whether it acquired regulatory sequences that lay near to its insertion site, or evolved new regulatory regions after the insertion of the transposed β-globin gene.

Whatever the origin of the regulatory sequences, their sequence divergence implies that they must have evolved either very rapidly, or independently in birds, monotremes and therians. The availability of sequences from other amniotes would make it possible to distinguish between these models by allowing the use of bioinformatics approaches to efficiently identify their regulatory regions and compare them across a broad range of species, revealing whether these regions are highly conserved or unique. This would

help clarify the origin of these regulatory sequences as well, for example, complete lack of conservation would mean that they originated independently.

## Future directions

This study has brought us a step closer to understand globin gene organization, function, regulation and evolution in amniotes. However, there is still more research needed to fill in the missing pieces of the puzzle.

First, to gain a broader insight into their gene content, arrangement, expression, function and evolution, sequence information is needed for the entire α- and β-globin clusters in non-avian reptiles such as snakes, tuatara, turtles and crocodiles. The next throughput in scientific revolution will be international moves to sequence genomes of 10,000 vertebrate species, approximately one for every vertebrate genus (Genome 10K Community of Scientists (G10KCOS, 2009)). With sufficient depth and complete assembly, one can obtain sequences covering the entire α- and β-globin clusters in many different species. Using the bioinformatics tools used in this study, all globin genes and their regulatory regions could be characterised in a range of amniotes. This would then help answer questions of the origin and evolution (and possible loss) of *GBY*, *HBW*, embryonic α- and β-like globin genes and/or any other novel globin gene.

It will also be necessary to study the expression of all globin genes throughout development using RT-PCR or micro-array techniques to further advance our understanding of the function of different globin genes in reptiles. Comparison of expression profiles with those in birds or mammals should furnish clues about the possible functions of these globins in non-erythroid cells, and the evolution of these functions over time in amniotes. Importantly, research into globins of more reptile species will facilitate phylogenetic studies. Exploring their true relationship with other globin genes, in particular, the β-globins, will provide answers to many evolutionary questions.

Secondly, even though my study shows that the origin of the amniote β-globin locus is by transposition of a copy of an ancient globin to a new site, it is not possible to distinguish between very different hypotheses to explain how the globin regulatory elements originated and how they evolved. It will be necessary to find the regulatory

elements of the platypus (and its one relative, echidna) β-globin locus, by searching further fields in the genome, using a bioinformatics approach or an experimental approach such as chromatin immunoprecipitation (ChIP), which determines the location of genomic transcription factor binding sites via immunoprecipitation of an endogenous transcription factor.

Understanding the origin and mechanism of these regulatory regions will be important in understanding how the cascade of regulatory elements ensure an equal expression of α- and β-globins in erythroid cells, as well as solving the mystery of their expression patterns in non-erythroid cells of the platypus. This search could be widened to the other monotreme species, the echidna, as well as to other reptiles, and would explicitly answer whether some regulatory elements were also transposed with the β-globin gene into the new region or not. If the regulatory regions of the echidna and platypus, non-avian reptiles and birds are similar, this would mean that regulatory elements were transposed with the β-globin gene, or were supplied at by neighbouring sequences at the transposition site, but evolved very rapidly in some species. Alternatively, if their regulatory regions show little similarity in sequences, locations and functions, this would suggest that the regulatory elements evolved independently after the transposition event.

Finally, we lack direct evidence that the α-β-*GBY* arrangement is ancestral. It is important to confirm the arrangement in amphibians to provide support to this study, perhaps by mapping these genes in *Xenopus tropicalis*. It would be extremely helpful to study other amphibians, and more frogs from different cohorts to clarify the order and timing of *HBK* duplication, the location of α-β-*GBY* on one or two chromosomes using FISH, and the expression pattern of the α- and β-like globin genes at different stages of development and in different tissues using RT-PCR or micro-array technologies. This information will then provide constructive comparisons with amniotes and help strengthen all hypotheses concerning the duplications and divergences that occurred at the stem of tetrapod evolution.

Globin genes provide of the best-studied systems for exploring genome structure, function and evolution. New data will advance the iterative process of discovery. More research into other species will help make our understanding of the evolution of α- and β-globin genes, their functions, clusters and regulatory regions much clearer.

# REFERENCES


Abbasi A, Wells RM, Brittain T, et al. (1988) Primary structure of the hemoglobins from Sphenodon (Sphenodon punctatus, Tuatara, Rynchocephalia). Evidence for the expression of alpha D-gene. Biol Chem Hoppe Seyler, 369: 755-764.

Aguileta G, Bielawski JP and Yang Z (2006a) Evolutionary rate variation among vertebrate beta globin genes: implications for dating gene family duplication events. Gene, 380: 21-29.

Aguileta G, Bielawski JP and Yang Z (2006b) Proposed standard nomenclature for the alpha- and beta-globin gene families. Genes Genet Syst, 81: 367-371.

Akiyama Y (Unpublished) TFSEARCH: Searching Transcription Factor Binding Sites. http://www.rwcp.or.jp/papia/.

Albitar M, Cash FE, Peschle C, et al. (1992) Developmental switch in the relative expression of the alpha 1- and alpha 2-globin genes in humans and in transgenic mice. Blood, 79: 2471-2474.

Albitar M, Peschle C and Liebhaber SA (1989) Theta, zeta, and epsilon globin messenger RNAs are expressed in adults. Blood, 74: 629-637.

Alev C, Shinmyozu K, McIntyre BA, et al. (2009) Genomic organization of zebra finch alpha and beta globin genes and their expression in primitive and definitive blood in comparison with globins in chicken. Dev Genes Evol.

Allen SJ, O'Donnell A, Alexander ND, et al. (1997) alpha+-Thalassemia protects children against disease caused by other infections as well as malaria. Proc Natl Acad Sci U S A, 94: 14736-14741.

Alsop AE, Miethke P, Rofe R, et al. (2005) Characterizing the chromosomes of the Australian model marsupial Macropus eugenii (tammar wallaby). Chromosome Res, 13: 627-636.

Banville D, Kay RM, Harris R, et al. (1983) The nucleotide sequence of the mRNA encoding a tadpole beta-globin polypeptide of Xenopus laevis. J Biol Chem, 258: 7924-7927.

Banville D and Williams JG (1985a) Developmental changes in the pattern of larval beta-globin gene expression in Xenopus laevis. Identification of two early larval beta-globin mRNA sequences. J Mol Biol, 184: 611-620.

Banville D and Williams JG (1985b) The pattern of expression of the Xenopus laevis tadpole alpha-globin genes and the amino acid sequence of the three major tadpole alpha-globin polypeptides. Nucleic Acids Res, 13: 5407-5421.

Bassimitci S, Yucel-Eroglu E and Akalar M (1996) Effects of thalassaemia major on components of the craniofacial complex. Br J Orthod, 23: 157-162.

Baumann R, Fischer J and Engelke M (1987) Functional properties of primitive and definitive red cells from chick embryo: oxygen-binding characteristics, pH and membrane potential, and response to hypoxia. J Exp Zool Suppl, 1: 227-238.

Bavisotto L, Kaushansky K, Lin N, et al. (1991) Antisense oligonucleotides from the stage-specific myeloid zinc finger gene MZF-1 inhibit granulopoiesis in vitro. J Exp Med, 174: 1097-1101.

Berenbrink M (2006) Evolution of vertebrate haemoglobins: Histidine side chains, specific buffer value and Bohr effect. Respir Physiol Neurobiol, 154: 165-184.

Bernard R, Little PFR, Annison G, et al. (1979) Structure of the human gammaG-gammaA-delta-beta-globin gene locus. Proc Natl Acad Sci U S A, 76: 4827-4831.

Bininda-Emonds OR, Cardillo M, Jones KE, et al. (2007) The delayed rise of present-day mammals. Nature, 446: 507-512.

Bisbee CA, Baker MA, Wilson AC, et al. (1977) Albumin phylogeny for clawed frogs (Xenopus). Science, 195: 785-787.

Blanchette M, Kent WJ, Riemer C, et al. (2004) Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res, 14: 708-715.

Briehl RW (1963) The Relation between the Oxygen Equilibrium and Aggregation of Subunits in Lamprey Hemoglobin. J.Mol.Chem., 238: 2361-2366.

Brittain T, Hofmann OM, Watmough NJ, et al. (1997) A two-state analysis of co-operative oxygen binding in the three human embryonic haemoglobins. Biochem J, 326 ( Pt 2): 299-303.

Brownlie A, Hersey C, Oates AC, et al. (2003) Characterization of embryonic globin genes of the zebrafish. Dev Biol, 255: 48-61.

Bruns GA and Ingram VM (1973) Erythropoiesis in the developing chick embryo. Dev Biol, 30: 455-459.

Bulger M, van Doorninck JH, Saitoh N, et al. (1999) Conservation of sequence and structure flanking the mouse and human beta-globin loci: the beta-globin genes are embedded within an array of odorant receptor genes. Proc Natl Acad Sci U S A, 96: 5129-5134.

Burge C and Karlin S (1997) Prediction of complete gene structures in human genomic DNA. J.Mol.Biol., 268: 78-94.

Burmester T, Ebner B, Weich B, et al. (2002) Cytoglobin: a novel globin type ubiquitously expressed in vertebrate tissues. Mol Biol Evol, 19: 416-421.

Burmester T, Gerlach F and Hankeln T (2007) Regulation and role of neuroglobin and cytoglobin under hypoxia. Adv Exp Med Biol, 618: 169-180.

Burmester T, Haberkamp M, Mitz S, et al. (2004) Neuroglobin and cytoglobin: genes, proteins and evolution. IUBMB Life, 56: 703-707.

Burmester T, Weich B, Reinhardt S, et al. (2000) A vertebrate globin expressed in the brain. Nature, 407: 520-523.

Cheng Y, Wu W, Kumar SA, et al. (2009) Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression. Genome Res, 19: 2172-2184.

Choi OR and Engel JD (1986) A 3' enhancer is required for temporal and tissue-specific transcriptional activation of the chicken adult beta-globin gene. Nature, 323: 731-734.

Coates ML (1975) Hemoglobin function in the vertebrates: an evolutionary model. J Mol Evol, 6: 285-307.

Cooper DW, VandeBerg JL, Griffiths ME, et al. (1973) Haemoglobin polymorphisms in the echidna, Tachyglossus aculeatus. Aust J Biol Sci, 26: 605-612.

Cooper SJ and Hope RM (1993) Evolution and expression of a beta-like globin gene of the Australian marsupial Sminthopsis crassicaudata. Proc Natl Acad Sci U S A, 90: 11777-11781.

Cooper SJ, Murphy R, Dolman G, et al. (1996) A molecular and evolutionary study of the beta-globin gene family of the Australian marsupial Sminthopsis crassicaudata. Mol Biol Evol, 13: 1012-1022.

Cooper SJ, Wheeler D, De Leo A, et al. (2006) The mammalian alphaD-globin gene lineage and a new model for the molecular evolution of alpha-globin gene clusters at the stem of the mammalian radiation. Mol Phylogenet Evol, 38: 439-448.

Cooper SJ, Wheeler D, Hope RM, et al. (2005) The alpha-globin gene family of an Australian marsupial, Macropus eugenii: the long evolutionary history of the theta-globin gene and its functional status in mammals. J Mol Evol, 60: 653-664.

Czelusniak J, Goodman M, Hewett-Emmett D, et al. (1982) Phylogenetic origins and adaptive evolution of avian and mammalian haemoglobin genes. Nature, 298: 297-300.

Dabrowski Z, Sano Martins IS, Tabarowski Z, et al. (2007) Haematopoiesis in snakes (Ophidia) in early postnatal development. Cell Tissue Res, 328: 291-299.

De Leo AA, Guedelha N, Toder R, et al. (1999) Comparative chromosome painting between marsupial orders: relationships with a 2n = 14 ancestral marsupial karyotype. Chromosome Res, 7: 509-517.

De Leo AA, Wheeler D, Lefevre C, et al. (2005) Sequencing and mapping hemoglobin gene clusters in the Australian model dasyurid marsupial Sminthopsis macroura. Cytogenet Genome Res, 108: 333-341.

de Souza PC and Bonilla-Rodriguez GO (2007) Fish hemoglobins. Braz J Med Biol Res, 40: 769-778.

Deakin JE, Koina E, Waters PD, et al. (2008) Physical map of two tammar wallaby chromosomes: a strategy for mapping in non-model mammals. Chromosome Res, 16: 1159-1175.

Dickerson RE and Geis I (1983) Hemoglobin. The Benjamin/Cummings Publishing Company, Inc., Menlo Park, Calif, 65-116

Dodgson JB, McCune KC, Rusling DJ, et al. (1981) Adult chicken alpha-globin genes alpha A and alpha D: no anemic shock alpha-globin exists in domestic chickens. Proc Natl Acad Sci U S A, 78: 5998-6002.

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics, 5: 113.

Elnitski L, Li J, Noguchi CT, et al. (2001) A negative cis-element regulates the level of enhancement by hypersensitive site 2 of the beta-globin locus control region. J Biol Chem, 276: 6289-6298.

Engel JD and Dodgson JB (1980) Analysis of the closely linked adult chicken alpha-globin genes in recombinant DNAs. Proc Natl Acad Sci U S A, 77: 2596-2600.

Ewens WJ, Griffiths RC, Ethier SN, et al. (1992) Statistical analysis of in situ hybridization data: derivation and use of the zmax test. Genomics, 12: 675-682.

Ezaz T, Moritz B, Waters P, et al. (2009) The ZW sex microchromosomes of an Australian dragon lizard share no homology with those of other reptiles or birds. Chromosome Res, 17: 965-973.

Fago A, Hundahl C, Malte H, et al. (2004) Functional properties of neuroglobin and cytoglobin. Insights into the ancestral physiological roles of globins. IUBMB Life, 56: 689-696.

Flavell RA, Kooter JM, De Boer E, et al. (1978) Analsyis of the beta-delta-globin gene loci in normal and Hb Lepore DNA: direct determination of gene linkage and intergene distance. Cell, 15: 25-41.

Flicek P, Aken BL, Ballester B, et al. (2010) Ensembl's 10th year. Nucleic Acids Res, 38: D557-562.

Flint J, Tufarelli C, Peden J, et al. (2001) Comparative genome analysis delimits a chromosomal domain and identifies key regulatory elements in the alpha globin cluster. Hum Mol Genet, 10: 371-382.

Forrester WC, Thompson C, Elder JT, et al. (1986) A developmentally stable chromatin structure in the human beta-globin gene cluster. Proc Natl Acad Sci U S A, 83: 1359-1363.

Fritsch EF, Lawn RM and Maniatis T (1980) Molecular cloning and characterization of the human beta-like globin gene cluster. Cell, 19: 959-972.

Fucharoen S and Winichagoon P (1997) Hemoglobinopathies in Southeast Asia: molecular biology and clinical medicine. Hemoglobin, 21: 299-319.

Fuchs C, Burmester T and Hankeln T (2006) The amphibian globin gene repertoire as revealed by the Xenopus genome. Cytogenet Genome Res, 112: 296-306.

Fushitani K, Higashiyama K, Moriyama EN, et al. (1996) The amino acid sequences of two alpha chains of hemoglobins from Komodo dragon Varanus komodoensis and phylogenetic relationships of amniotes. Mol Biol Evol, 13: 1039-1043.

G10KCOS (2009) Genome 10K: a proposal to obtain whole-genome sequence for 10,000 vertebrate species. J Hered, 100: 659-674.

Gaboli M, Kotsi PA, Gurrieri C, et al. (2001) Mzf1 controls cell proliferation and tumorigenesis. Genes Dev, 15: 1625-1630.

Garner C, Dew TK, Sherwood R, et al. (2003) Heterocellular hereditary persistence of fetal haemoglobin affects the haematological parameters of beta-thalassaemia. Br J Haemat, 123: 353-358.

Garner KJ and Lingrel JB (1988) Structural organization of the beta-globin locus of B-haplotype sheep. Mol Biol Evol, 5: 134-140.

Garner KJ and Lingrel JB (1989) A comparison of the beta A-and beta B-globin gene clusters of sheep. J Mol Evol, 28: 175-184.

Gillemans N, McMorrow T, Tewari R, et al. (2003) Functional and comparative analysis of globin loci in pufferfish and humans. Blood, 101: 2842-2849.

Goh SH, Lee YT, Bhanu NV, et al. (2005) A newly discovered human alpha-globin gene. Blood, 106: 1466-1472.

Goodman M, Czelusniak J, Koop BF, et al. (1987) Globins: a case study in molecular phylogeny. Cold Spring Harb Symp Quant Biol, 52: 875-890.

Goodman M, Koop BF, Czelusniak J, et al. (1984) The eta-globin gene. Its long evolutionary history in the beta-globin gene family of mammals. J Mol Biol, 180: 803-823.

Goodman M, Moore GW, Barnabas J, et al. (1974) The phylogeny of human globin genes investigated by the maximum parsimony method. J Mol Evol, 3: 1-48.

Goodman M, Moore GW and Matsuda G (1975) Darwinian evolution in the genealogy of haemoglobin. Nature, 253: 603-608.

Goodman M, Pedwaydon J, Czelusniak J, et al. (1988) An evolutionary tree for invertebrate globin sequences. J Mol Evol, 27: 236-249.

Gorr TA, Mable BK and Kleinschmidt T (1998) Phylogenetic analysis of reptilian hemoglobins: trees, rates, and divergences. J Mol Evol, 47: 471-485.

Green P (2007) 2x genomes--does depth matter? Genome Res, 17: 1547-1549.

Grosveld F, Antoniou M, van Assendelft GB, et al. (1987a) The regulation of expression of human beta-globin genes. Prog Clin Biol Res, 251: 133-144.

Grosveld F, van Assendelft GB, Greaves DR, et al. (1987b) Position-independent, high-level expression of the human beta-globin gene in transgenic mice. Cell, 51: 975-985.

Guindon S and Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol, 52: 696-704.

Hardison R (1998) Hemoglobins from bacteria to man: evolution of different patterns of gene expression. J Exp Biol, 201: 1099-1117.

Hardison R and Miller W (1993) Use of long sequence alignments to study the evolution and regulation of mammalian globin gene clusters. Mol Biol Evol, 10: 73-102.

Hardison R, Slightom JL, Gumucio DL, et al. (1997) Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. Gene, 205: 73-94.

Hardison RC (2005) Globin Genes: Evolution. Encylcopedia of Life Sciences. John Wiley & Sons, Ltd: Chichester. http://www.els.net/ [doi:10.1038/npg.els.0005134],

Hardison RC (2008) Globin genes on the move. J Biol, 7: 35.

Hardison RC, Sawada I, Cheng JF, et al. (1986) A previously undetected pseudogene in the human alpha globin gene cluster. Nucleic Acids Res, 14: 1903-1911.

Hedges SB and Kumar S (Eds.) (2009) The Timetree of Life, Oxford, Oxford University Press.

Hedges SB and Vidal N (2009) Lizards, snakes, and amphisbaenians (Squamata). In: Hedges SB and Kumar S, eds. The Timetree of Life. Oxford University Press, pp 383-389

Heinemeyer T, Wingender E, Reuter I, et al. (1998) Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. Nucleic Acids Res, 26: 362-367.

Hesse JE, Nickol JM, Lieber MR, et al. (1986) Regulated gene expression in transfected primary chicken erythrocytes. Proc Natl Acad Sci U S A, 83: 4312-4316.

Hiebl I, Schneeganss D and Braunitzer G (1986) High-altitude respiration of birds. The primary structures of the alpha D-chains of the Bar-headed Goose (Anser indicus), the Greylag Goose(Anser anser) and the Canada Goose (Branta canadensis). Biol Chem Hoppe Seyler, 367: 591-599.

Hiebl I, Schneeganss D, Grimm F, et al. (1987) High altitude respiration of birds. The primary structures of the major and minor hemoglobin component of adult European black vulture (Aegypius monachus, Aegypiinae). Biol Chem Hoppe Seyler, 368: 11-18.

Hiebl I, Weber RE, Schneeganss D, et al. (1988) High-altitude respiration of birds. Structural adaptations in the major and minor hemoglobin components of adult Ruppell's Griffon (Gyps rueppellii, Aegypiinae): a new molecular pattern for hypoxic tolerance. Biol Chem Hoppe Seyler, 369: 217-232.

Higgs DR, Vernimmen D and Wood B (2008) Long-range regulation of alpha-globin gene expression. Adv Genet, 61: 143-173.

Higgs DR, Wood WG, Jarman AP, et al. (1990) A major positive regulatory region located far upstream of the human alpha-globin gene locus. Genes Dev, 4: 1588-1601.

Hoffmann FG, Opazo JC and Storz JF (2008) Rapid rates of lineage-specific gene duplication and deletion in the alpha-globin gene family. Mol Biol Evol, 25: 591-602.

Hoffmann FG, Opazo JC and Storz JF (2010a) Gene cooption and convergent evolution of oxygen transport hemoglobins in jawed and jawless vertebrates. Proc Natl Acad Sci U S A, 107: 14274-14279.

Hoffmann FG and Storz JF (2007) The {alpha}D-Globin Gene Originated via Duplication of an Embryonic {alpha}-Like Globin Gene in the Ancestor of Tetrapod Vertebrates. Mol Biol Evol, 24: 1982-1990.

Hoffmann FG, Storz JF, Gorr TA, et al. (2010b) Lineage-Specific Patterns of Functional Diversification in the {alpha}- and {beta}-Globin Gene Families of Tetrapod Vertebrates. Mol Biol Evol.

Holland RA and Gooley AA, 2nd (1997) Characterization of the embryonic globin chains of the marsupial Tammar wallaby, Macropus eugenii. Eur J Biochem, 248: 864-871.

Holland RA, Gooley AA and Hope RM (1998) Embryonic globins of the marsupial the Tammar Wallaby (Macropus eugenii): bird like and mammal like. Clin Exp Pharmacol Physiol, 25: 740-744.

Hosbach HA, Wyler T and Weber R (1983) The Xenopus laevis globin gene family: chromosomal arrangement and gene structure. Cell, 32: 45-53.

Hromas R, Collins SJ, Hickstein D, et al. (1991) A retinoic acid-responsive human zinc finger gene, MZF-1, preferentially expressed in myeloid cells. J Biol Chem, 266: 14183-14187.

Hsu SL, Marks J, Shaw JP, et al. (1988) Structure and expression of the human theta 1 globin gene. Nature, 331: 94-96.

Hughes JR, Cheng JF, Ventress N, et al. (2005) Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences. Proc Natl Acad Sci U S A, 102: 9830-9835.

Inoue M, Takahashi K, Niide O, et al. (2005) LDOC1, a novel MZF-1-interacting protein, induces apoptosis. FEBS Lett, 579: 604-608.

Isaacks RE, Harkness DR and Witham PR (1978) Relationship between the major phosphorylated metabolic intermediates and oxygen affinity of whole blood in the loggerhead (Caretta caretta) and the Green Sea Turtle (Chelonia mydas) during development. Dev Biol, 62: 344-353.

Jeffreys AJ, Wilson V, Wood D, et al. (1980) Linkage of adult alpha- and beta-globin genes in X. laevis and gene duplication by tetraploidization. Cell, 21: 555-564.

Kanehisa M, Goto S, Hattori M, et al. (2006) From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res, 34: D354-357.

Kassahn KS, Dang VT, Wilkins SJ, et al. (2009) Evolution of gene function and regulatory control after whole-genome duplication: Comparative analyses in vertebrates. Genome Res.

Katoh K and Miyata T (2002) Cyclostome hemoglobins are possibly paralogous to gnathostome hemoglobins. J Mol Evol, 55: 246-249.

Kawada N, Kristensen DB, Asahina K, et al. (2001) Characterization of a stellate cell activation-associated protein (STAP) with peroxidase activity found in rat hepatic stellate cells. J Biol Chem, 276: 25318-25323.

Kent WJ, Sugnet CW, Furey TS, et al. (2002) The human genome browser at UCSC. Genome Res, 12: 996-1006.

Koop BF and Goodman M (1988) Evolutionary and developmental aspects of two hemoglobin beta-chain genes (epsilon M and beta M) of opossum. Proc Natl Acad Sci U S A, 85: 3893-3897.

Kugelstadt D, Haberkamp M, Hankeln T, et al. (2004) Neuroglobin, cytoglobin, and a novel, eye-specific globin from chicken. Biochem Biophys Res Commun, 325: 719-725.

Labie D and Elion J (1996) Sequence polymorphisms of potential functional relevance in the beta-globin gene locus. Hemoglobin, 20: 85-101.

Lander ES, Linton LM, Birren B, et al. (2001) Initial sequencing and analysis of the human genome. Nature, 409: 860-921.

Lanfranchi G, Pallavicini A, Laveder P, et al. (1994) Ancestral hemoglobin switching in lampreys. Dev Biol, 164: 402-408.

Larkin MA, Blackshields G, Brown NP, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics, 23: 2947-2948.

Lauer J, Shen CK and Maniatis T (1980) The chromosomal arrangement of human alpha-like globin genes: sequence homology and alpha-globin gene deletions. Cell, 20: 119-130.

Lee MH, Shroff R, Cooper SJ, et al. (1999) Evolution and molecular characterization of a beta-globin gene from the Australian Echidna Tachyglossus aculeatus (Monotremata). Mol Phylogenet Evol, 12: 205-214.

Leung S, Proudfoot NJ and Whitelaw E (1987) The gene for theta-globin is transcribed in human fetal erythroid tissues. Nature, 329: 551-554.

Li SL and Riggs A (1970) The amino acid sequence of hemoglobin V from the lamprey, Petromyzon marinus. J Biol Chem, 245: 6149-6169.

Manwell C (1960) Comparative physiology: blood pigments. Annu Rev Physiol, 22: 191-244.

Maruyama K, Ishikawa Y, Yasumasu S, et al. (2007) Globin gene enhancer activity of a DNase-I hypersensitive site-40 homolog in medaka, Oryzias latipes. Zoolog Sci, 24: 997-1004.

Maruyama K, Yasumasu S and Iuchi I (2002) Characterization and expression of embryonic and adult globins of the teleost Oryzias latipes (medaka). J Biochem, 132: 581-589.

Maruyama K, Yasumasu S, Naruse K, et al. (2004) Genomic organization and developmental expression of globin genes in the teleost Oryzias latipes. Gene, 335: 89-100.

Mason MM, Lee E, Westphal H, et al. (1995) Expression of the chicken beta-globin gene cluster in mice: correct developmental expression and distributed control. Mol Cell Biol, 15: 407-414.

McCutcheon FH (1947) Specific oxygen affinity of hemoglobin in elasmobranchs and turtles. J Cell Physiol, 29: 333-344.

Melo MB, Bordin S, Duarte AS, et al. (2003) Molecular characterization of hemoglobin alpha-D chains from Geochelone carbonaria and Geochelone denticulata land turtles. Comp Biochem Physiol B Biochem Mol Biol, 134: 389-395.

Merlet-Benichou C (1975) [Oxygen transport by the fetal blood]. Poumon Coeur, 31: 197-203.

Mockenhaupt FP, Ehrhardt S, Gellert S, et al. (2004) Alpha(+)-thalassemia protects African children from severe malaria. Blood, 104: 2003-2006.

Morris JF, Hromas R and Rauscher FJ, 3rd (1994) Characterization of the DNA-binding properties of the myeloid zinc finger protein MZF1: two independent DNA-binding domains recognize two DNA consensus sequences with a common G-rich core. Mol Cell Biol, 14: 1786-1795.

Mudduluru G, Vajkoczy P and Allgayer H (2010) Myeloid zinc finger 1 induces migration, invasion, and in vivo metastasis through Axl gene expression in solid cancer. Mol Cancer Res, 8: 159-169.

Opazo JC, Hoffmann FG and Storz JF (2008a) Differential loss of embryonic globin genes during the radiation of placental mammals. Proc Natl Acad Sci U S A, 105: 12950-12955.

Opazo JC, Hoffmann FG and Storz JF (2008b) Genomic evidence for independent origins of beta-like globin genes in monotremes and therian mammals. Proc Natl Acad Sci U S A, 105: 1590-1595.

Orkin SH (1978) The duplicated human alpha globin genes lie close together in cellular DNA. Proc Natl Acad Sci U S A, 75: 5950-5954.

Ovcharenko I, Loots GG, Giardine BM, et al. (2005) Mulan: multiple-sequence local alignment and visualization for studying function and evolution. Genome Res, 15: 184-194.

Patel VS, Cooper SJ, Deakin JE, et al. (2008) Platypus globin genes and flanking loci suggest a new insertional model for beta-globin evolution in birds and mammals. BMC Biol, 6: 34.

Patient RK, Elkington JA, Kay RM, et al. (1980) Internal organization of the major adult alpha- and beta-globin genes of X. laevis. Cell, 21: 565-573.

Perutz MF (1984) Species adaptation in a protein molecule. Adv Protein Chem, 36: 213-244.

Peterson FC, Hayes PL, Waltner JK, et al. (2006) Structure of the SCAN domain from the tumor suppressor protein MZF1. J Mol Biol, 363: 137-147.

Pisano E, Cocca E, Mazzei F, et al. (2003) Mapping of alpha- and beta-globin genes on Antarctic fish chromosomes by fluorescence in-situ hybridization. Chromosome Res, 11: 633-640.

Pough FH (1969) Environmental adaptations in the blood of lizards. Comp Biochem Physiol, 31: 885-901.

Pough FH (1977) Ontogenetic change in molecular and functional properties of blood of garter snakes, Thamnophis sirtalis. J Exp Zool, 201: 47-55.

Proudfoot NJ and Maniatis T (1980) The structure of a human alpha-globin pseudogene and its relationship to alpha-globin gene duplication. Cell, 21: 537-544.

Quinn NL, Boroevich KA, Lubieniecki KP, et al. (2010) Genomic organization and evolution of the Atlantic salmon hemoglobin repertoire. BMC Genomics, 11: 539.

Reitman M and Felsenfeld G (1990) Developmental regulation of topoisomerase II sites and DNase I-hypersensitive sites in the chicken beta-globin locus. Mol Cell Biol, 10: 2774-2786.

Reitman M, Grasso JA, Blumenthal R, et al. (1993a) Primary sequence, evolution, and repetitive elements of the Gallus gallus (chicken) beta-globin cluster. Genomics, 18: 616-626.

Reitman M, Lee E and Westphal H (1995) Function of the upstream hypersensitive sites of the chicken beta-globin gene cluster in mice. Nucleic Acids Res, 23: 1790-1794.

Reitman M, Lee E, Westphal H, et al. (1993b) An enhancer/locus control region is not sufficient to open chromatin. Mol Cell Biol, 13: 3990-3998.

Rens W, Grutzner F, O'brien PC, et al. (2004) Resolution and evolution of the duck-billed platypus karyotype with an X1Y1X2Y2X3Y3X4Y4X5Y5 male sex chromosome constitution. Proc Natl Acad Sci U S A, 101: 16257-16261.

Roesner A, Fuchs C, Hankeln T, et al. (2005) A globin gene of ancient evolutionary origin in lower vertebrates: evidence for two distinct globin families in animals. Mol Biol Evol, 22: 12-20.

Rozen S and Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol, 132: 365-386.

Rucknagel KP and Braunitzer G (1988) Hemoglobins of reptiles. The primary structure of the major and minor hemoglobin component of adult Western Painted Turtle (Chrysemys picta bellii). Biol Chem Hoppe Seyler, 369: 123-131.

Rumen NM and Love WE (1963) The Six Hemoglobins of the Sea Lamprey (Petromyzon Marinus). Arch Biochem Biophys, 103: 24-35.

Sayers EW, Barrett T, Benson DA, et al. (2010) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res, 39: D38-51.

Schalekamp M and Van Goor D (1984) Haemoglobin and globin synthesis in the isolated primitive and definitive erythroid cells of chicken embryos. Evidence for a non-clonal mechanism at the haemoglobin switch. J Embryol Exp Morphol, 84: 125-148.

Schimenti JC and Duncan CH (1985) Structure and organization of the bovine beta-globin genes. Mol Biol Evol, 2: 514-525.

Schwartz S, Zhang Z, Frazer KA, et al. (2000) PipMaker--a web server for aligning two genomic DNA sequences. Genome Res, 10: 577-586.

Shedlock AM and Edwards SV (2009) Amniotes (Amniota). In: Hedges SB and Kumar S, eds. The Timetree of Life. Oxford, Oxford University Press, 375-379

Sinclair AH and Graves JA (1991) Gene mapping in marsupials: detection of an ancient autosomal gene cluster. Genomics, 9: 581-586.

Stamatoyannopoulos G (2005) Control of globin gene expression during development and erythroid differentiation. Exp Hematol, 33: 259-271.

Stoeckelhuber M (1992) Sequenzermittlung von Schlangenhaemoglobinen. Primaerstrucktur von Globinketten der Indigonatter (Drymarchon corais erebenus) und des Tigerpython (Python molurus bivittatus). Diploma thesis, University of Munich.

Tanaka Y, Eishi Y and Morris B (1988) Splenic hemopoiesis of the platypus (Ornithorhynchus anatinus): evidence of primary hemopoiesis in the spleen of a primitive mammal. Am J Anat, 181: 401-405.

Taylor TD, Noguchi H, Totoki Y, et al. (2006) Human chromosome 11 DNA sequence and analysis including novel gene identification. Nature, 440: 497-500.

Thompson EO, Fisher WK and Whittaker RG (1973) Studies on monotreme proteins. 3. Amino acid sequence of the alpha- and beta-globin chains of the minor haemoglobin from the echidna, Tachyglossus aculeatus aculeatus. Aust J Biol Sci, 26: 1327-1335.

Townes TM, Fitzgerald MC and Lingrel JB (1984) Triplication of a four-gene set during evolution of the goat beta-globin locus produced three genes now expressed differentially during development. Proc Natl Acad Sci U S A, 81: 6589-6593.

Trandafir F, Hoogewijs D, Altieri F, et al. (2007) Neuroglobin and cytoglobin as potential enzyme or substrate. Gene, 398: 103-113.

Trent JT, 3rd and Hargrove MS (2002) A ubiquitously expressed human hexacoordinate hemoglobin. J Biol Chem, 277: 19538-19545.

Vyas P, Vickers MA, Picketts DJ, et al. (1995) Conservation of position and sequence of a novel, widely expressed gene containing the major human alpha-globin regulatory element. Genomics, 29: 679-689.

Wang H, Zhang Y, Cheng Y, et al. (2006) Experimental validation of predicted mammalian erythroid cis-regulatory modules. Genome Res, 16: 1480-1492.

Weatherall DJ (1991) The structure, organization, and regulation of human genes, Oxford, Oxford University Press.

Wells RMG (1999) Evolution of haemoglobin function: Molecular adaptations to environment. Clinical and Experimental Pharmacology and Physiology, 26: 591-595.

Wells RMG and Brennan SO (1992) The detection and importance of functionally abnormal haemoglobins. N.Z. Med. J., 105: 329-330.

Wheeler D, Hope R, Cooper SB, et al. (2001) An orphaned mammalian beta-globin gene of ancient evolutionary origin. Proc Natl Acad Sci U S A, 98: 1101-1106.

Wheeler D, Hope RM, Cooper SJ, et al. (2004) Linkage of the beta-like omega-globin gene to alpha-like globin genes in an Australian marsupial supports the chromosome duplication model for separation of globin gene clusters. J Mol Evol, 58: 642-652.

Whittaker RG, Fisher WK and Thompson EO (1972) Studies on monotreme proteins. I. Amino acid sequence of the beta-chain of haemoglobin from the echidna, Tachyglossus aculeatus aculeatus. Aust J Biol Sci, 25: 989-1004.

Whittaker RG and Thompson EO (1974) Studies on monotreme proteins. V. Amino acid sequence of the alpha-chain of haemoglobin from the platypus, *Ornithorhynchus anantinus*. Aust J Biol Sci, 27: 591-605.

Whittaker RG and Thompson EO (1975) Studies on monotreme proteins. VI. Amino acid sequence of the beta-chain of haemoglobin from the platypus, Ornithorhynchus anatinus. Aust J Biol Sci, 28: 353-365.

Wittenberg JB and Wittenberg BA (2003) Myoglobin function reassessed. J Exp Biol, 206: 2011-2020.

Yeh RF, Lim LP and Burge CB (2001) Computational inference of homologous gene structures in the human genome. Genome Res, 11: 803-816.

# APPENDIX 1: PHYSICAL MAP OF TWO TAMMAR WALLABY CHROMOSOMES: A STRATEGY FOR MAPPING IN NON-MODEL MAMMALS

***Authors:***

Deakin, J.E., Koina, E., Waters, D., Doherty, R., **Patel, V.S.**, Delbridge, M.L., et al.

***Current status of paper (circle as appropriate):***

Planned/ In Preparation/ Submitted/ Under Revision/ Accepted/ Published

***Date paper accepted for publication or anticipated date of publication:***

2 September 2008

***Name of journal:***

Chromosome Research **16**: 1159-1175

***Extent to which research is your own:***

Although my contribution was minor, the time taken to perform experiment on five genes (*CCKBR, ILK, HBB, FTSJ1, RRM1*) was considerable. I designed probes for isolation of BAC clones containing these genes and then mapped them onto the tammar wallaby chromosomes using FISH.

***Your contribution to writing the paper:***

I was involved in editing the manuscript as a co-author and contributed to a picture in Figure 5.

***Comments:***

My role was to isolate and map the β-globin cluster and its flanking genes in the tammar wallaby and the information was used to generate the physical map of the wallaby chromosome 5.

Chromosome Research (2008) 16:1159–1175
DOI: 10.1007/s10577-008-1266-y

© Springer 2008

# Physical map of two tammar wallaby chromosomes: A strategy for mapping in non-model mammals

Janine E. Deakin[1]*, Edda Koina[1], Paul D. Waters[1], Ruth Doherty[1], Vidushi S. Patel[1], Margaret L. Delbridge[1], Bianca Dobson[1], James Fong[1], Yanqiu Hu[2], Cecilia van den Hurk[1], Andrew J. Pask[2], Geoff Shaw[2], Carly Smith[1], Katherine Thompson[1], Matthew J. Wakefield[3], Hongshi Yu[2], Marilyn B. Renfree[2] & Jennifer A. Marshall Graves[1]

[1]*ARC Centre of Excellence for Kangaroo Genomics, Research School of Biological Sciences, The Australian National University, Canberra ACT 0200, Australia; Tel: +61-2-6125-2101; Fax: +61-2-6125-489; E-mail: Janine.Deakin@anu.edu.au;* [2]*ARC Centre of Excellence for Kangaroo Genomics, Department of Zoology, University of Melbourne, Victoria 3010, Australia;* [3]*The Walter and Eliza Hall Institute, 1G Royal Parade, Parkville, Victoria 3050, Australia*

* *Correspondence*

Received 30 July 2008. Received in revised form and accepted for publication by Fengtang Yang 2 September 2008

*Key words:* comparative mapping, genomics, marsupial, X chromosome evolution

## Abstract

Marsupials are especially valuable for comparative genomic studies of mammals. Two distantly related model marsupials have been sequenced: the South American opossum (*Monodelphis domestica*) and the tammar wallaby (*Macropus eugenii*), which last shared a common ancestor about 70 Mya. The six-fold opossum genome sequence has been assembled and assigned to chromosomes with the help of a cytogenetic map. A good cytogenetic map will be even more essential for assembly and anchoring of the two-fold wallaby genome. As a start to generating a physical map of gene locations on wallaby chromosomes, we focused on two chromosomes sharing homology with the human X, wallaby chromosomes X and 5. We devised an efficient strategy for mapping large conserved synteny blocks in non-model mammals, and applied this to generate dense maps of the X and 'neo-X' regions and to determine the arrangement of large conserved synteny blocks on chromosome 5. Comparisons between the wallaby and opossum chromosome maps revealed many rearrangements, highlighting the need for comparative gene mapping between South American and Australian marsupials. Frequent rearrangement of the X, along with the absence of a marsupial *XIST* gene, suggests that inactivation of the marsupial X chromosome does not depend on a whole-chromosome repression by a control locus.

## Abbreviations

| | | | |
|---|---|---|---|
| BAC | bacterial artificial chromosome | DAPI | 4′,6-diamidino-2-phenylindole dihydrochloride |
| BLASTN | basic local alignment search tool nucleotide | dUTP | 2′-deoxyuridine 5′-triphosphate |
| bp | base pair(s) | FISH | fluorescence in-situ hybridization |
| CCD | charge-coupled device | GGA | *Gallus gallus* |
| cDNA | complementary DNA | HSA | *Homo sapiens* |

## Electronic supplementary material

The online version of this article (doi: 10.1007/s10577-008-1266-y) contains supplementary material, which is available to authorized users.

| | |
|---|---|
| kb | kilo base pairs |
| LB | Luria broth |
| Mb | mega base pairs |
| MDO | *Monodelphis domestica* |
| MHC | major histocompatibility complex |
| Mya | million years ago |
| OR | olfactory receptor |
| PCR | polymerase chain reaction |
| RISH | radioactive in-situ hybridization |
| SSC | standard sodium citrate |
| XAR | X added region |
| XCI | X chromosome inactivation |
| XCR | X conserved region |
| XIC | X inactivation centre |

## Introduction

The inclusion of marsupials in comparative genomic studies has significantly contributed to our understanding of the evolution of genes and chromosomes, the identification of novel genes and regulatory sequences, and the function and evolution of complex genetic regulatory systems (Wakefield & Graves 2003). Marsupials last shared a common ancestor with eutherian mammals between 145 (Bininda-Emonds *et al.* 2007) and 180 million years ago (Mya) (Woodburne *et al.* 2003) and occupy the important phylogenetic position between bird–mammal divergence (310 Mya) and the eutherian radiation (105 Mya).

To further our understanding of mammalian genome evolution, structure and function, two marsupial genomes have been sequenced: the South American grey short-tailed opossum (*Monodelphis domestica)* and the Australian model kangaroo, the tammar wallaby (*Macropus eugenii*). The opossum has been used predominantly as a model species for biomedical research (Samollow 2006), whereas the wallaby has been the model marsupial species for classic studies of marsupial physiology and reproduction (Tyndale-Biscoe & Renfree 1987; Tyndale-Biscoe 2005) and has been used for genetic studies for over 30 years. Comparing opossum and wallaby genomes will reveal many important and informative differences, as well as identify marsupial-specific genome features, such as novel protein-coding genes or marsupial-specific non-coding elements. Since Australian and American marsupials diverged around 70 Mya (Kirsch *et al.* 1997), this comparison will be similar in evolutionary terms to that between human and mouse.

The opossum genome was sequenced by the Broad Institute (USA) on average six times over (i.e. to a depth of six-fold) and the sequenced scaffolds were assigned to chromosomes (Duke *et al.* 2007; Mikkelsen *et al.* 2007). The wallaby genome has recently been sequenced jointly by the Australian Genome Research Facility (AGRF, Melbourne) and Baylor College of Medicine Human Genome Sequencing Center (USA) to provide a two-fold coverage of the genome (National Human Genome Institute 2008). With less sequence coverage, the wallaby assembly will consist of many smaller contigs and many more gaps than the opossum assembly. As there is currently only a rudimentary physical map for the wallaby (Alsop *et al.* 2005), anchoring each of these small contigs to chromosomes would be an extremely labour-intensive process. However, the assignment of sequence to chromosomes can at least determine the extent of rearrangement that has taken place between these two species, as well as illuminate the evolutionary history of marsupial chromosomes.

Marsupial genomes are characteristically packaged into just a few very large chromosomes, which show an astounding level of conservation even between distantly related species (Rens *et al.* 2003; Rofe & Hayman 1985). Chromosome painting studies suggest that all marsupial karyotypic diversity is the result of the rearrangement of only 19 evolutionary blocks (Rens *et al.* 2003). Although chromosome painting identifies regions of homology, it does not indicate gene order or internal rearrangements, which can be exposed only by the finer detail provided by gene mapping.

Knowledge of the large regions of homology between opossum and wallaby revealed by chromosome painting, combined with sequence data from the opossum genome project (Mikkelsen *et al.* 2007), makes it possible to predict which genes will be on each wallaby chromosome. This provides a starting point for gene mapping, particularly when targeting regions or chromosomes of interest.

We are particularly interested in the organization, function and evolution of the mammalian X chromosome. Gene mapping and chromosome painting between marsupials and eutherians showed that the human X chromosome can be divided into an ancient region shared with the marsupial X, and a recently added region that is autosomal in marsupials (Glas *et al.* 1999; Graves 1995; Wilcox *et al.* 1996). Genes on the long arm and pericentric region of the human X also map to the wallaby X chromosome, defining an ancient conserved region that was part of the

mammalian X chromosome since marsupials and eutherians last shared a common ancestor between 145 and 180 Mya.

Genes on the rest of the short arm of the human X are located on chromosome 5 in wallaby, defining a region that was added to the X in eutherians since their divergence from marsupials. This region is formally a neo-X. Some early reports assigned a few genes from this added region to wallaby chromosome 1, suggesting two additions (Spencer *et al.* 1991a). Opossum sequencing confirmed the division of the human X into a conserved region (on the X in the opossum), and a recently added region (split between opossum chromosomes 4 and 7) (Mikkelsen *et al.* 2007), and mapping in chicken confirms that the X conserved region (XCR) and X added region (XAR) were originally present as at least two separate evolutionary blocks (Kohn *et al.* 2004). In eutherians, most of the genes located in the conserved region of the X undergo X chromosome inactivation (XCI), but many genes on the recently added region escape inactivation (Carrel & Willard 2005).

XCI is a unique regulatory mechanism that compensates for the difference in gene dosage between XX females and XY males in mice and humans (Heard & Disteche 2006). Eutherian XCI is a multi-stage, chromosome-wide process (Gartler *et al.* 1985) under the control of the *XIST* (X inactive specific transcript) locus within the X inactivation centre (XIC) (Brown *et al.* 1991), which transcriptionally silences genes (Graves & Gartler 1986) on one X chromosome in female somatic cells. X inactivation also occurs in marsupials, but our understanding of marsupial XCI is very limited. Marsupial XCI studies, performed on only five genes, reveal differences in activation status between marsupial species and tissues and even between genes (reviewed in Cooper *et al.* 1993). Marsupial XCI differs markedly from the random and relatively complete inactivation of eutherians. Marsupials preferentially silence the paternal X chromosome (Sharman 1971) and inactivation is incomplete and tissue specific (Cooper *et al.* 1993). It is therefore important to study more X-borne loci in a single species. The absence of an *XIST* gene in marsupials implies that control of inactivation is quite different from that in eutherians (Davidow *et al.* 2007; Duret *et al.* 2006; Hore *et al.* 2007; Shevchenko *et al.* 2007).

The previously published wallaby physical map had a mere 18 genes assigned to the X chromosome and only 16 to chromosome 5 (Alsop *et al.* 2005). A new strategy has now been devised to make the mapping of genes onto wallaby chromosomes more efficient. We can use chromosome painting data to predict which genes are present on wallaby chromosomes, then use sequence generated from the wallaby genome project to design specific overgo probes for screening the wallaby BAC library for those specific genes. Single and multicolour mapping of BACs to tammar chromosomes by FISH is extremely efficient.

In this study we aimed to fill in the map of wallaby chromosomes X and 5, define the limits of the region homologous to the human neo-X, and construct a human–wallaby comparative map of the region. Here we report the construction of dense physical maps of the wallaby X and the recently added neo-X region, as well as the anchoring of large evolutionary conserved blocks on chromosome 5.

## Materials and methods

### *PCR-generated probes*

Prior to the wallaby genome sequencing project and the use of a more efficient strategy using pools of overgos, several genes were screened for by using PCR-generated probes. Primers for these genes were designed from sequence conserved between opossum and human to amplify tammar wallaby orthologues of *AR* (AF100638), *CDX2*, *CTPS2*, *PHKA2*, *POLA1*, *REPS2*, *RNF12* (designed from BAC sequence CR385048), *SMS*, *TEX11* and *WNT4* (AY940685). All primers and their annealing temperatures are listed in Table 1 and probe sequences were submitted to GenBank if not already present in the database (EU743940–EU743944). These primers were used to amplify probes and/or to confirm identity of isolated BACs. *TEX11* primers were used only to confirm that BACs isolated by screening with *TEX11* overgos were positive. A BAC previously mapped as an X chromosome anchor BAC (MeVIA_72C1) by Alsop *et al.* (2005) was found to contain the *RBMX* gene using *RBMX* primers on the BAC. Products were either cloned into the TA TOPO Cloning kit (Invitrogen Corp., Carlsbad, CA, USA) and sequenced with vector primers, or directly sequenced using PCR primers. Sequencing reactions were performed by AGRF (Brisbane, Australia).

*Table 1.* Primers used to generate PCR probes for library screening and/or to test positive BACs

| Gene | Forward primer | Reverse primer | Product size (bp) | Annealing temperature (°C) | Product |
|---|---|---|---|---|---|
| AR | gcccggaagctgaagaaact | agggcctcggcccacttgac | 284 | 60 | exon 4 |
| CDX2 | catgagcatgtayccyagc | cctggyrctgagscttgc | 807 | 58.5 | exons 1–5 (cDNA) |
| CTPS2 | ggtgaagtgtttgtattaaatgatgg | ataatcaccatgcctttctttaattg | 153 | 58 | exon 2 |
| KAL1 | cgttcttcaagcayttccag | aaggcttgtagtgatgdggatg | 1839 | 53 | exons 1–14 (cDNA) |
| PHKA2 | tctgtcttgcagagaggtgcag | ctggtagaaccccactggaa | 162 | 58 | exon 29 |
| POLA1 | acacctcagatagctccaccac | gagtcacaggaccttcctttctta | 604 | 58 | exons 6–7 (plus intron) |
| RBMX | cagagattatgctccaccac | tagatatcacttcggctgct | ~1100 | 60 | exons 7–9 (plus introns) |
| REPS2 | cagtgatgtggactgtgatgg | agattcaggcagtgggtagc | 100 | 58 | exon 9 |
| RNF12 | cagatgatgtgtctaatggtgactc | tagaacgtcttgcagatggc | 225 | 55 | exon 4 |
| SMS | gcagagagtgacttggcgta | tgaccatctttggtttcaattt | 139 | 58 | exon 6 |
| TEX11 | ctgaagccctgcactggtat | ccttgtccagctgcttcaga | 120 | 59 | exon 17 |
| WNT4[a] | gacggtggaactgctcgactctg | attctagaggccgaggcggccgaca | 952 | 60 | exon 1–polyA tail (cDNA) |

[a]From Yu *et al.* (2006).

*Overgo design*

To design overgo primers for each gene of interest, opossum orthologues from the Ensembl gene build (MonDom5) were used to BLAST search the wallaby genome sequence (*Macropus eugenii* WGS) in the trace archives (http://www.ncbi.nlm.nih.gov). Over-go primers were designed by the Overgo Maker program, which is available for download (Washington University Genome Sequencing Center 2008) using the wallaby orthologous sequence as the input sequence (Supplementary Tables S1 and S2). To check for specificity, the resulting 40 bp probe sequences were screened against the opossum genome assembly and wallaby trace archives using BLASTN. Overgo specificity was judged to be good if the entire 40 bp hit only those wallaby trace sequences from which the overgo was designed. Overgos with hits to numerous wallaby traces or positions in the opossum genome were excluded in order to avoid detection of paralogous copies of genes.

*Library screening*

High-density filters from the tammar wallaby BAC library (Me_KBa; Arizona Genomics Institute, Tucson, AZ, USA) were screened with pools of radioactively labelled PCR probes or overgos as described in (Deakin *et al.* 2007). Resulting positive BACs were further screened with individual probes using dot blots of either purified BAC DNA spotted onto Hybond N+ membrane (GE Healthcare UK Ltd, Little Chalfont, Buckinghamshire, UK) or from BAC cultures spotted on Hybond N+ filters placed on LB plates containing 12 μg/ml chloramphenicol and then grown overnight at 37°C. The colonies on the Hybond N+ filters were then lysed, denatured and neutralized using the manufacturer's recommended protocol, and fixed to the membrane by placing on Whatman 3MM chromatography paper (Whatman International Ltd, Maidstone, Kent, UK) soaked in 0.4 M NaOH for 20 min. Dot blots were hybridized overnight using the same protocol as for library screening.

Two overlapping BACs (MeVIA_22L1 and MeVIA_35G12) containing *OPN1MW*, *TKTL1* and *TEX11* were isolated using cDNA probes as described by Wakefield *et al.* (2008).

*Direct sequencing*

A portion of BACs (41 out of a total of 272) were subjected to direct sequencing, using overgos as sequencing primers, to confirm the accuracy of the library screening followed by dot blotting technique. Among the BACs sequenced were the eight BACs that did not map to predicted location on chromosome 5 but instead mapped to chromosome 6. Four hundred nanograms of BAC DNA was used in a sequencing reaction containing 4 μl BigDye Terminator version 3.1 (Applied Biosystems, Foster City, CA, USA) and 5 pmol primer. Cycling conditions for sequencing reactions were: 95°C 5 min followed by 99 cycles of 96°C for 10 s, 50°C for 10 s and 60°C for 4 min. Reactions were precipitated and sent to AGRF (Brisbane, Australia) for capillary separation.

*Fluorescence in-situ hybridization (FISH)*

DNA (~1 μg) from each BAC clone isolated was labelled by nick translation with digoxygenin-11-dUTP or biotin-16-dUTP (Roche Diagnostics, Basel, Switzerland), SpectrumRed dUTP, SpectrumOrange dUTP or SpectrumGreen dUTP (Abbott Molecular Inc., Des Plaines, IL, USA), and hybridized to male wallaby metaphase chromosomes following the protocol described in Alsop *et al.* (2005). Digoxygenin and biotin fluorescent signals were detected following the protocol described previously (Alsop *et al.* 2005). Slides hybridized with probes labelled directly with SpectrumRed, SpectrumOrange or SpectrumGreen were washed at 60°C in 0.4× SSC with 0.3% (v/v) Tween 20 for 2 min, followed by a 5 s to 1 min wash at room temperature in 2× SSC with 0.1% (v/v) Tween 20 to remove unbound probe. Chromosomes were counterstained with DAPI in Vectashield (Vector Laboratories Inc., Burlingame, CA, USA) and a Zeiss Axioplan2 epifluorescence microscope was used to visualize fluorescent signals. Images of fluorescent signals and DAPI staining were captured on a SPOT RT Monochrome CCD (charge-coupled device) camera (Diagnostic Instruments Inc., Sterling Heights, MI, USA) and merged using IP Lab imaging software (Scanalytics Inc, Fairfax, VA, USA).

*Comparative and virtual map construction*

Opossum, human and chicken orthologues of protein-coding genes lying between (and including) mapped genes in the wallaby were downloaded from Ensembl v49 with the biomart tool (http://www.ensembl.org/Multi/martview) using the opossum as the reference genome. Gene order for the virtual map was determined based on the order of physically mapped genes. Only those genes with human orthologues and belonging to the conserved synteny block in human were included in the virtual map.

## Results

We used a combination of chromosome painting data and wallaby genome sequence to predict which genes would be found on our target chromosomes, X and 5. We then isolated BACs bearing these target genes and confirmed the validity of the BAC library screening followed by dot blot approach by obtaining sequence from 41 of the total 272 BACs isolated, using the overgos as sequencing primers. All 41 BACs were determined to contain the gene of interest. BACs were mapped using fluorescence in-situ hybridization (FISH), thus generating cytogenetic maps of these two chromosomes. Gene mapping localizations for chromosome 5 was used to construct a virtual map, by inferring the location of genes based on the opossum assembly. By comparing the virtual map to the opossum, human and chicken genomes, we were able to determine the extent of conserved synteny between these species and deduce rearrangements between wallaby and opossum, and between marsupials and humans.

*Physical map of the wallaby X chromosome*

Previous chromosome painting experiments showed that the wallaby X has overall homology at the sequence level to the opossum X (Rens *et al.* 2003) and to the long arm and pericentric region of the human X (Glas *et al.* 1999). We therefore predicted that all genes on the opossum X, and on the human XCR, would lie on the wallaby X. We chose at least one gene from each band of the human X chromosome. We obtained homologous sequence from the wallaby trace archives to design overgos from each of these genes, which we used to screen the wallaby BAC library. Ten genes were screened for prior to the availability of wallaby genome sequence. PCR primers (Table 1) were used to amplify homologous wallaby sequence, which we then used to screen the BAC library, and/or to confirm the identity of BACs.

A total of 47 new genes was mapped by FISH to the long arm of the wallaby X in this study, as well as 26 from previous studies (Table 2), making a total of 73 genes now mapped to the wallaby X. We determined gene order by performing two- or three-colour FISH experiments (see Figure 1A–E for examples).

Among the 47 genes we mapped were five that had previously been mapped by radioactive in-situ hybridization (RISH) using human cDNA probes: *AR* (Spencer *et al.* 1991c), *F9* (Spencer *et al.* 1991b), *GATA1* (Wilcox *et al.* 1996), *GLA*, *OPN1MW* (Spencer *et al.* 1991c), and two assigned to the X by segregation from rodent-tammar somatic cell hybrids: *PGK1*, *HPRT1* (Graves *et al.* 1979). Our mapping provides more reliable and precise localizations of genes previously mapped by RISH, and

*Table 2.* Genes previously mapped onto the X chromosome. Genes with a corrected location have been mapped by FISH in the present study

| Gene(s) | Location | Reference | Corrected location |
|---|---|---|---|
| **FISH** | | | |
| *AFF2, ARHGEF6, AGTR2, FMR1, OPHN1, RPS6KA6, ATRX* | Xq | Delbridge *et al.* (2008)<br>Pask *et al.* (2000) | |
| *CACNA1F, CCDC22, FOXP3, GPR173, HSD17B10, HUWE1, IQSEC2, JARID1C, LMO6, SYP, PPP1R3F, RIBC1* | Xq3 | Delbridge *et al.* in preparation | |
| *TMEM187, IRAK, MEPC2* | Xq1 | Koina *et al.* in preparation | |
| *G6PD* (*IKBKG* - CR956372) | Xq1 | Koina & Graves (2005) | |
| *PLP1* | Xq3 | Koina & Graves (2006) | |
| *SLC16A2* | Xq3 | Koina *et al.* (2005) | |
| **RISH** | | | |
| *ALAS2* | Xp | Wilcox *et al.* (1996) | |
| *AR* | Xq (proximal) | Spencer *et al.* (1991c) | |
| *ARAF* | Xq (distal) | Wilcox *et al.* (1996) | |
| BGN | Xq (proximal) | Wilcox *et al.* (1996) | |
| F8 | Xq (proximal) | Spencer *et al.* (1991b) | |
| F9 | Xq (proximal) | Spencer *et al.* (1991b) | |
| *GATA1* | Xp | Wilcox *et al.* (1996) | Xq3 |
| *GLA* | Xq (proximal) | Spencer *et al.* (1991b) | Xq3 |
| *OPN1LW* | Xq (distal) | Spencer *et al.* (1991b) | Xq1 |
| *PLP1* | Xq (proximal) | Spencer *et al.* (1991b) | Xq3 |
| *SLC10A3* | Xq (proximal) | Spencer *et al.* (1991b) | |
| *SOX3* | Xq | Foster & Graves (1994) | |
| *UBL4A* | Xq | Spencer *et al.* (1991b) | |
| **Somatic cell hybrids** | | | |
| *PGK* | X | Graves *et al.* (1979) | Xq3 |
| *HPRT1* | X | Graves *et al.* (1979) | Xq2 |

localization of the genes assigned by somatic cell hybridization.

We also corrected the assignment of genes from the human Xp region that had previously been localized to two different autosomes by RISH. *SYN1* and *TBC1D251 (OATL)* from human Xp11 were previously mapped to wallaby 1p and 3/4p respectively (Spencer *et al.* 1991a), but we found that BACs containing these genes both localized to Xq with other genes corresponding to the same region in human, including *GATA1* which had been incorrectly mapped to wallaby Xp by RISH (Wilcox *et al.* 1996).

A comparison between the gene arrangement in wallaby and opossum shows that many rearrangements have occurred on the marsupial X (Figure 2). For instance, the genes *PSMD10*, *TBC1D8B* and *IL1RAPL2* are located together in the wallaby X but map to three separate positions in opossum X. These three genes also lie together in human Xq22.3, implying that rearrangements occurred in the opossum. Comparison of gene order in the wallaby and human X reveals many rearrangements (Figure 2). Of particular note is the separation of genes from the human *XIST* region Xq13, with *STARD8* assigned to proximal Xq and *SLC16A2*, *RNF12*, *OGT*, *TEX11* and *PGK1* located distally in wallaby.

### *Chromosome 5 physical and virtual maps*

Chromosome painting previously showed that wallaby chromosome 5 shares large regions of homology with opossum chromosomes 4 and 7 (Rens *et al.* 2003). Using this information as a guide and comparing the opossum and human genomes, we were able to densely map a region corresponding to part of the recently added region of the human X chromosome, as well as to identify 15 large evolutionarily conserved blocks (represented as differently coloured blocks in Figure 3B). Genes at the endpoints of these blocks were isolated from the BAC library and physically mapped using FISH (Figure 3). The genes mapped and their corresponding BACs are

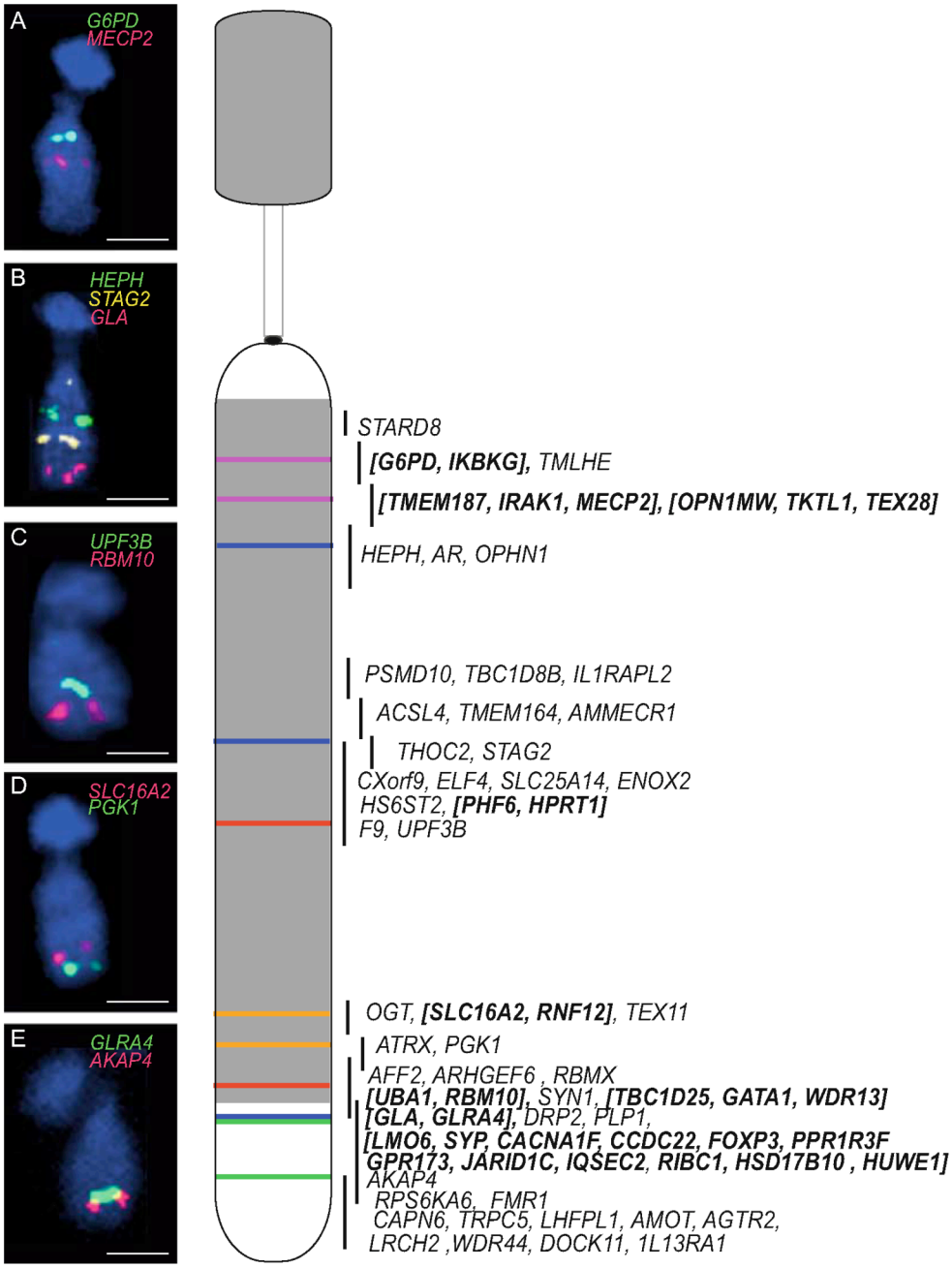



*Figure 1.* Physical map of the wallaby X chromosome. The DAPI banding pattern is indicated in grey. Genes found within one BAC or overlapping BACs are shown in bold and brackets. Coloured stripes on chromosomes indicate the location of genes shown in inserts (A, pink; B, blue; C, red; D, orange; E, green). Scale bars represent 1 μm.

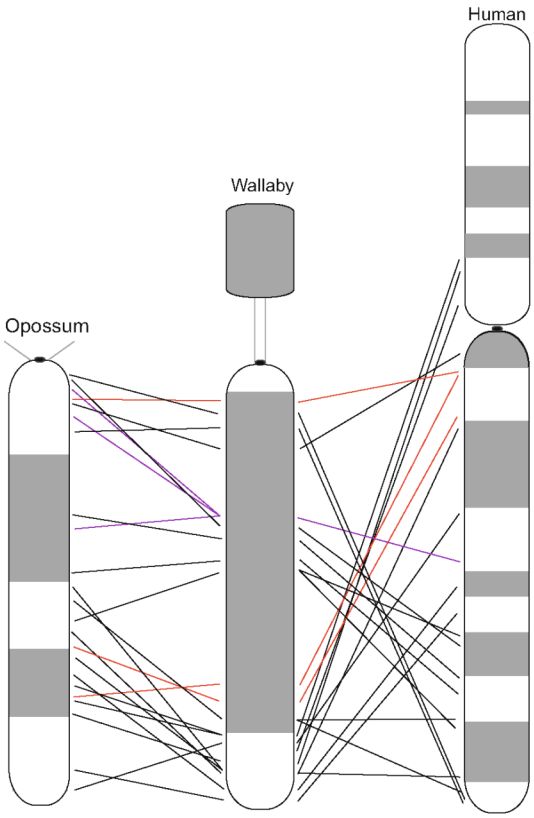



*Figure 2*. Comparison of the wallaby X with opossum and human X chromosomes. Genes from Xq13 (XIC-containing band) are indicated in red and genes *PSMD10*, *RBC1D8B* and *IL1RAPL2* are in purple to highlight the rearrangement of these genes in opossum.

listed in Supplementary Table S2. Where possible, we determined the orientation of these conserved blocks. We mapped 125 genes to chromosome 5. With the 16 genes mapped to chromosome 5 by FISH in previous studies (see Table 3), this brings the total to 141. Extrapolating from the opossum genome assembly, this allowed us to infer the location of approximately 2320 additional protein-coding genes (Supplementary Table S3).

A comparative map was constructed by comparing gene blocks on wallaby chromosome 5 with opossum, human and chicken. There are at least four large-scale rearrangements that have occurred between wallaby and opossum as indicated in Figure 3. Block 1 is inverted compared to opossum chromosome 4, block 2 is located on a different chromosome in opossum (chromosome 7), and two inversion events have taken place on the long arm. Based on the opossum assembly, the breakpoints for these inversions were determined to be in the 4 Mb region between *TECTA* and *BACE1* and in the 3 Mb region between *SPG20* and *FREM2*. *TECTA* and *BACE1* are part of the same block with the same gene order in opossum and humans, whereas in chicken the order of these two genes with respect to flanking markers is different. The breakpoint between *SPG20* and *FREM2* is unique to the wallaby, as chicken, opossum and human all show the same gene order.

Chromosome 5 contains blocks of genes from nine different human chromosomes. It is clear from Figure 3 that blocks of genes now on human chromosomes 2, 15, 21 and X are on one chromosome in both the wallaby and chicken, indicating that these genes were together in the common ancestor of birds and mammals and became fragmented in eutherians.

Of particular interest was to explore the evolutionarily conserved block corresponding to genes from the human X chromosome Xp11.2 –22.1, which corresponds to the fusion point between the X and neo-X. This region was extensively mapped by localizing orthologues for 52 genes. In some instances, conservation of local gene order was confirmed by finding genes adjacent in opossum and human either located on the same BAC clone or on overlapping BACs. Previous mapping studies using RISH assigned four genes from this region of the human X (*CYBB*, *DMD*, *MAOA*, and *ZFX*) to wallaby 5p (Table 3) but *POLA1* to wallaby 1q (Sinclair *et al.* 1988; Spencer *et al.* 1991a). We confirmed that *CYBB*, *DMD*, *MAOA*, and *ZFX* lie on wallaby 5p, and mapped *POLA1* to the same region, making it

*Figure 3*. (**A**) Physical map of tammar wallaby chromosome 5. DAPI bands are indicated in grey. Genes contained within a BAC or overlapping BACs are indicated in bold and brackets. (**B**) Comparative map comparing wallaby chromosome 5 with opossum (MDO), human (HSA) and chicken (GGA). The orientation of genes within each block is shown by a diagonal line indicating direct or inverse collinearity. The absence of a line means that the orientation could not be determined by FISH mapping. A block with two colours indicates that the block consists of genes from two different chromosomes but the order cannot be ascertained. (**C**) Two-colour FISH revealing the break between *BACE1* and *TECTA*. (**D**) Two-colour FISH revealing the break between *SPG20* and *FREM2*. Scale bars represent 1 μm.

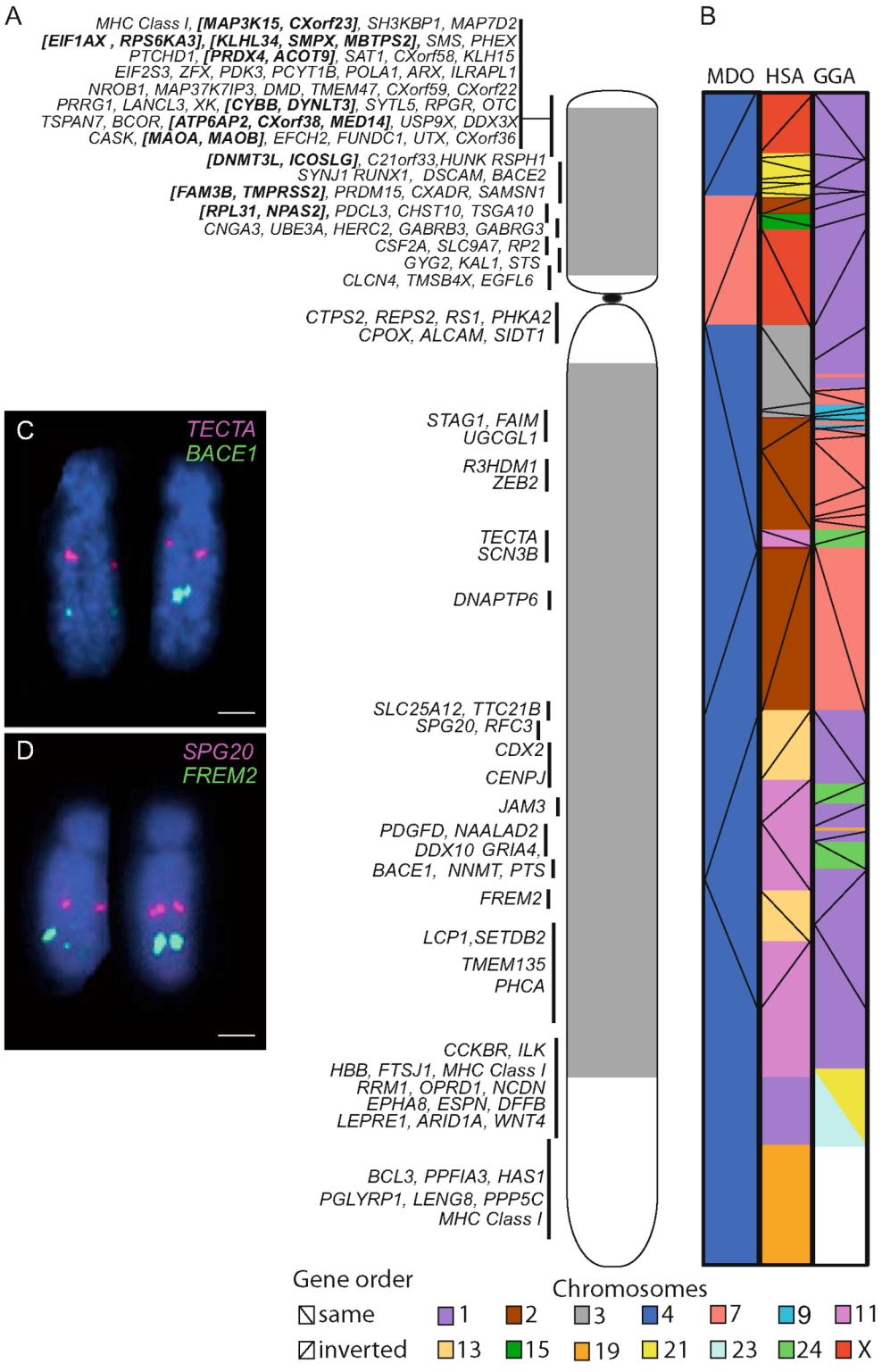

A
B
MDO HSA GGA
MHC Class I, [MAP3K15, CXorf23], SH3KBP1, MAP7D2
[EIF1AX , RPS6KA3], [KLHL34, SMPX, MBTPS2], SMS, PHEX
PTCHD1, [PRDX4, ACOT9], SAT1, CXorf58, KLH15
EIF2S3, ZFX, PDK3, PCYT1B, POLA1, ARX, ILRAPL1
NROB1, MAP37K7IP3, DMD, TMEM47, CXorf59, CXorf22
PRRG1, LANCL3, XK, [CYBB, DYNLT3], SYTL5, RPGR, OTC
TSPAN7, BCOR, [ATP6AP2, CXorf38, MED14], USP9X, DDX3X
CASK, [MAOA, MAOB], EFCH2, FUNDC1, UTX, CXorf36
[DNMT3L, ICOSLG], C21orf33,HUNK RSPH1
SYNJ1 RUNX1, DSCAM, BACE2
[FAM3B, TMPRSS2], PRDM15, CXADR, SAMSN1
[RPL31, NPAS2], PDCL3, CHST10, TSGA10
CNGA3, UBE3A, HERC2, GABRB3, GABRG3
CSF2A, SLC9A7, RP2
GYG2, KAL1, STS
CLCN4, TMSB4X, EGFL6
CTPS2, REPS2, RS1, PHKA2
CPOX, ALCAM, SIDT1
STAG1, FAIM
UGCGL1
R3HDM1
ZEB2
TECTA
SCN3B
DNAPTP6
SLC25A12, TTC21B
SPG20, RFC3
CDX2
CENPJ
JAM3
PDGFD, NAALAD2
DDX10 GRIA4,
BACE1, NNMT, PTS
FREM2
LCP1,SETDB2
TMEM135
PHCA
CCKBR, ILK
HBB, FTSJ1, MHC Class I
RRM1, OPRD1, NCDN
EPHA8, ESPN, DFFB
LEPRE1, ARID1A, WNT4
BCL3, PPFIA3, HAS1
PGLYRP1, LENG8, PPP5C
MHC Class I
C
TECTA
BACE1
D
SPG20
FREM2
Gene order
same
inverted
Chromosomes
1 2 3 4 7 9 11
13 15 19 21 23 24 X

*Table 3.* Genes previously assigned to chromosome 5

| Chromosome 5 | Location | Reference |
|---|---|---|
| **FISH** | | |
| *CNGA3, UBE3A, HERC2, GABRB3* | 5p | Rapkins *et al.* (2006) |
| *CSF2RA, STS* | 5p | Toder & Graves (1998) |
| *EIF2S3* | 5p | Delbridge & Graves (2004) |
| MHC Class I genes | 5p3, 5q2, 5q3 | Deakin *et al.* (2007) |
| *NROB1*[a] *(DAX1)* | 5p | Pask *et al.* (1997) |
| *TMSB4X* | 5p | Waters *et al.* (2003) |
| *TSPAN7, ARX, RPS6KA3* | 5p | Delbridge *et al.* (2008) |
| *ARID1A*[a] | 5q | Waters *et al.* (2001) |
| **RISH** | | |
| *CYBB, DMD, MAOA, ZFX* | 5p | Sinclair *et al.* (1988) |
| *PHDA1* | 5p | Fitzgerald *et al.* (1993) |
| *AMELX* | 5q | Watson *et al.* (1992) |

[a]Mapped to 5q using a BAC clone as part of the present study.

part of the large conserved block of human Xp genes located on wallaby 5p.

The block of human X genes located on opossum chromosome 7 were found to map to the pericentric region of wallaby chromosome 5, and were flanked by genes from human chromosomes 3, 15 and 2. A total of 23 genes corresponding to this region on opossum chromosome 7 were localized in wallaby in order to determine the endpoints, which are *RPL31* (HSA 2q11.2) and *CPOX* (HSA 3q12), and the approximate position of the centromere, which lies between *EGFL6* on 5p and *CTPS2* on 5q (Figure 4). Genes flanking this 62 Mb region in opossum map to wallaby 6q (Figure 4), except for one flanking gene, *MINA*, which is located at the distal end of 6q.

## Discussion

### *Mapping strategy and applications*

Chromosome painting has been a valuable technique for determining global chromosome homologies between species. However, this technique cannot detect important internal rearrangements. Gene mapping gives a more detailed and accurate picture, providing us with a better understanding of the evolutionary history of chromosomes.

Prior to our starting this study, there were only 18 genes assigned to the wallaby X and 16 assigned to wallaby chromosome 5 (Alsop *et al.* 2005). These genes were mapped mostly by RISH using heterologous probes (often human cDNAs) or by FISH using wallaby genomic clones isolated from lambda libraries with either homologous or heterologous probes. Mapping using short heterologous probes suffers from the complication of detecting processed pseudogenes, and RISH relies on statistical analysis to give only an approximate location, Not surprisingly, we found that some of these early map locations were in error.

We developed an efficient approach for generating a physical map of target chromosomes X and 5 in the wallaby. The availability of the wallaby genome trace sequences, overgos and BAC libraries has had a dramatic impact on the efficiency of map generation. We now have 73 and 141 genes assigned to the wallaby chromosomes X and 5 respectively. Additionally, our strategy of identifying large blocks with conserved synteny and mapping the ends of these blocks has allowed us to extrapolate from the opossum genome assembly and infer the location of an additional 2320 protein-coding genes onto a virtual map of chromosome 5.

This approach can be generalized to any other non-model species for which limited sequence data are available. It could be particularly useful for mammal species for which a two-fold sequencing coverage is being generated, where mapping data are available for closely related species. In addition, it can be used to generate virtual maps of reptile species by reference to the chicken genome and comparative painting across other birds and reptiles (Shetty *et al.* 1999).

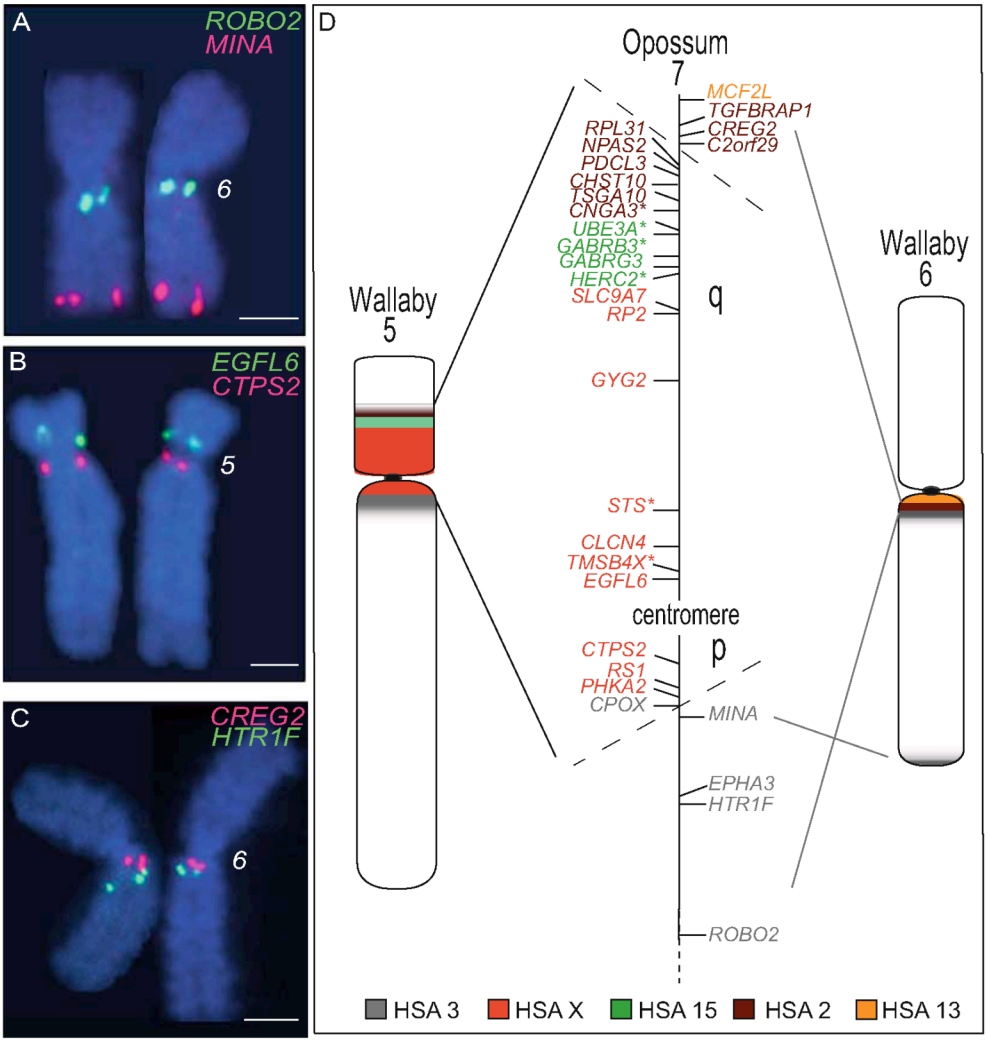



*Figure 4*. Details of region of homology on wallaby chromosome 5 with opossum chromosome 7. (**A**) Genes *ROBO2* and *MINA* from opossum chromosome 7 are located at different ends of wallaby chromosome 6. (**B**) *CLCN4* and *CTPS2* flank the centromere on wallaby chromosome 5. (**C**) *CREG2* and *HTR1F*, which flank the 62 Mb region in opossum, are adjacent on wallaby chromosome 6. (**D**) Diagram of region of homology on wallaby 5 compared to opossum 7 (only genes mapped in wallaby are listed). Opossum chromosome 7 is inverted to show the wallaby gene arrangement. Gene names are colour-coded based on their location in human.

Our mapping data are a valuable resource, being the first step towards anchoring wallaby genome sequence to chromosomes. It would be too costly and labour-intensive to assign to chromosomes each of the small contigs likely to arise from the assembly of the two-fold genome sequence coverage. In particular, the virtual map of wallaby chromosome 5 makes it possible to assign the assembly to chromosomes without the need to map every contig. Although the virtual map will be a great aid to genome assembly, it is important to note that it can only act as a guide, as some genes may be missing from the opossum genome assembly that are present in the wallaby and vice versa, or there may be small-scale rear-

rangements not detected by mapping just a few genes from these blocks. For example, MHC Class I genes have been assigned to three different locations on wallaby chromosome 5 (Deakin *et al.* 2007) but no orthologues of these genes are found on opossum chromosome 4.

### *The X chromosome, rearrangements and X inactivation*

The eutherian X chromosome is extremely conserved in gene content, perhaps because of selection against rearrangements that would disrupt the whole X chromosome dosage compensation system (Ohno 1967). The X in eutherians (with the exception of mice) has remained relatively unchanged, suggesting that the need for *XIST* transcripts to spread across the chromosome from the X inactivation centre (XIC) has resulted in a selection against structural rearrangements. It is therefore particularly interesting to examine the extent of this conservation in marsupials, which have an incomplete and tissue-specific inactivation system (Cooper *et al.* 1993).

Mikkelsen *et al.* (2007) showed that the opossum X has experienced several major rearrangements since divergence from eutherians. We now demonstrate many rearrangements between opossum and wallaby X chromosomes that were undetectable by chromosome painting. The extent of rearrangement of the X (Figure 2) which occurred since the wallaby and opossum last shared a common ancestor implies that in the marsupial X, gene order has not been highly conserved, suggesting that marsupial X chromosome inactivation does not depend on gene position.

This observation is consistent with the contention that, in marsupials, *XIST* does not coordinate X inactivation. Genes from human Xq13, flanking the region which contains *XIST* in humans, map far apart on the X chromosomes of wallaby and opossum Xq (Davidow *et al.* 2007, Hore *et al.* 2007, Shevchenko *et al.* 2007). The disruption of the region homologous to the eutherian X-inactivation centre in opossum and the failure to find any evidence of an *XIST* orthologue in opossum and platypus strongly suggest that *XIST* arose after the marsupial/eutherian divergence (Duret *et al.* 2006, Davidow *et al.* 2007, Hore *et al.* 2007; Shevchenko *et al.* 2007). The absence of *XIST* raises many questions regarding the control of inactivation in marsupials (Hore *et al.* 2007).

Much less is known about X inactivation in marsupials, compared with eutherians. However, with the large number of BACs now available for the wallaby X, the activity of many more X-borne genes can be studied via RNA-FISH, a technique that detects RNA at the site of transcription. This technique has already been used successfully to show the inactivation status of the gene *SLC16A2* in the wallaby (Koina *et al.* 2005). An activity map of the wallaby X would help distinguish whether marsupial XCI is controlled locally on a gene-by-gene basis or by spreading control of XCI over domains on the wallaby X, as observed in the human X.

### *Evolutionary history of genes on wallaby chromosome 5*

The generation of a virtual map of chromosome 5 from physical mapping data has made it possible to determine the extent to which synteny is conserved between distantly related species, and reveal some of the evolutionary history of the segments making up this chromosome. The evolutionary history of chromosomal segments can have important implications for the function, expression or control of genes within the region. For instance, revealing the evolutionary history of a chromosomal segment on chromosome 5 containing some of the genes corresponding to the imprinted region on human chromosome 15 responsible for Prader–Willi and Angelman syndromes also revealed that this region was assembled only after the divergence of marsupials and eutherians, and is not imprinted in marsupials (Rapkins *et al.* 2006).

Chromosome painting showed that wallaby chromosome 5 consists of two segments (referred to as C11 and C12) that are conserved among marsupials. C11 spans the entire short arm to just below the centromere on the long arm and C12 makes up the remainder of 5q (Rens *et al.* 2003). In the opossum, C11 is on part of the short arm and the long arm of opossum chromosome 4, and C12 constitutes chromosome 7. Our gene mapping data reveal the orientation of C12 and internal rearrangements within C11, and also shows that C12 does not encompass the entire short arm of chromosome 5 but is restricted to the pericentric region. Svartman & Vianna-Morgante (1998) suggested that evolutionary conserved blocks C10 (6q in wallaby), C11 and C12

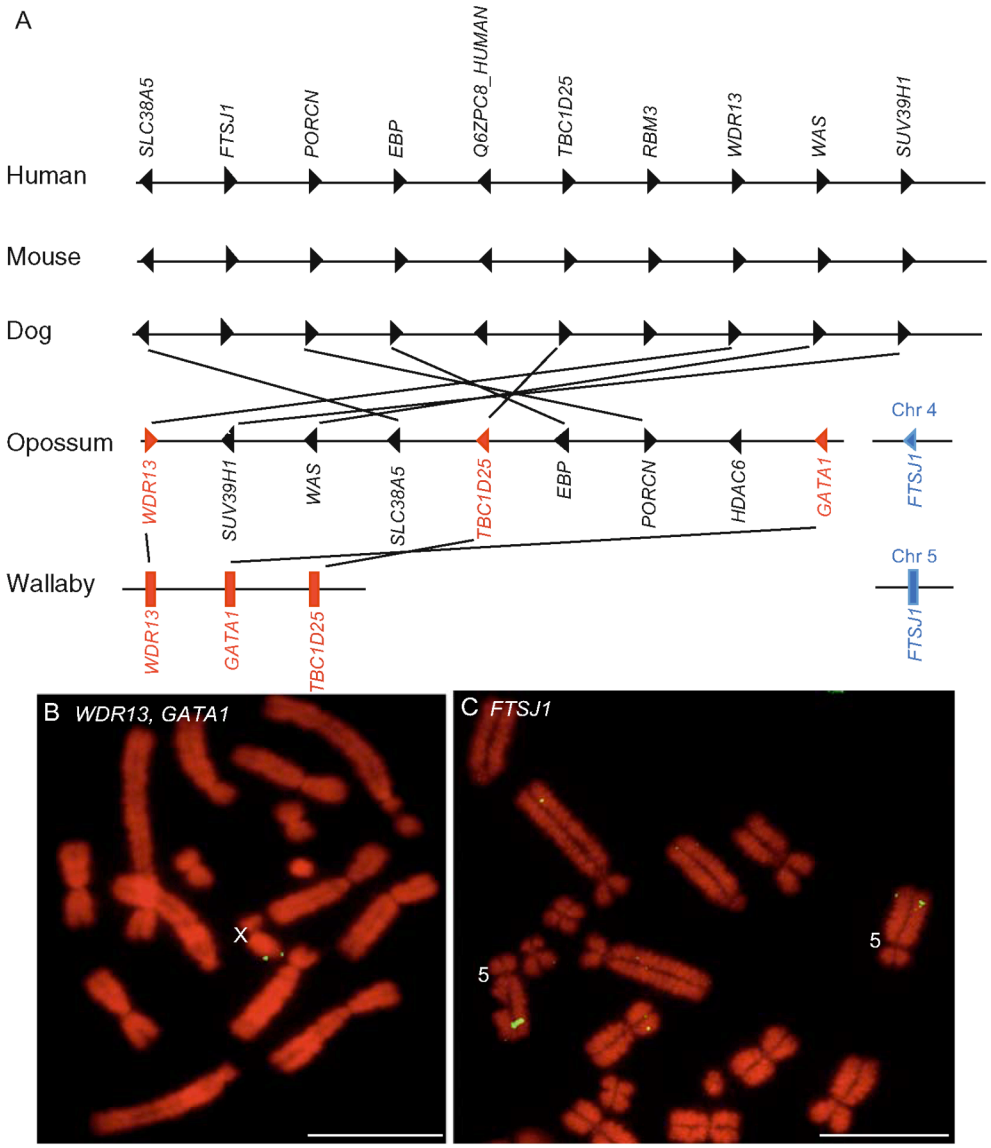



*Figure 5.* (**A**) Organization of the region surrounding *FTSJ1* in eutherians and marsupials. Gene order and orientation (indicated by arrows) are conserved among eutherians but genes have been completely rearranged in opossum and wallaby. *FTSJ1* is located on opossum chromosome 4 flanked by olfactory receptor genes and all other genes are on the X. (**B**) A BAC containing *WDR13* and *GATA1* maps to the X chromosome in wallaby and (**C**) *FTSJ1* is on chromosome 5. Scale bars represent 10 μm.

were on separate chromosomes in the ancestral marsupial. In contrast, combining information from chromosome painting (Rens *et al.* 2003) with genome sequence data (Ensembl 49) and our gene mapping data, it appears that in the common ancestor of birds and mammals all three segments were part of the same chromosome, which since underwent fission and fusion events in some marsupial lineages and remain together in the potoroo and species with a 2n=14 karyotype (Rens *et al.* 2003).

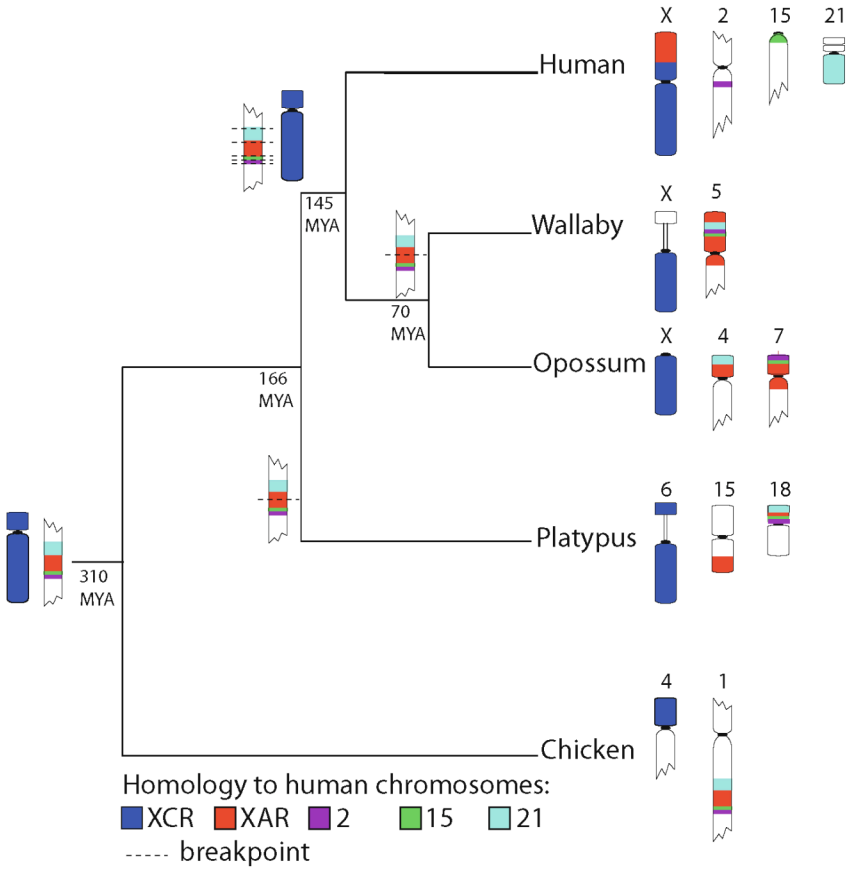



*Figure 6.* Refining the breakpoint between the X conserved (blue) and X added region (red). The XCR has remained as one block since prior to avian/mammal divergence. Genes making up the XAR region reside adjacent to genes found on human chromosomes 2 (purple), 15 (green) and 21 (light blue) on chicken chromosome 1. These genes have undergone different fission (indicated by broken lines) and subsequent fusion events in the three mammalian lineages. Assignment of platypus orthologues of human XCR, XAR and chromosome 21 genes is taken from Veyrunes *et al.* (2008) and assignment of orthologues of human chromosome 2 and 15 genes is taken from Edwards *et al.* (2007).

At first glance, wallaby 5p and the pericentric region of 5q appear to have retained an ancestral arrangement, since these genes are located on the same chromosome in chicken but are split over two chromosomes in opossum and four in human. However, a closer comparison of wallaby 5 and opossum chromosome 7 with chicken shows that the opossum gene order is more similar to chicken than is wallaby and, hence, represents the ancestral marsupial gene arrangement. Especially notable in wallaby is the separation of *MINA* from the other flanking genes to end up at the distal end of 6q (see Figure 3A). In the opossum, *MINA* is adjacent to a cluster of olfactory receptors (ORs). It is apparent from examining this region in other vertebrates that it has acquired olfactory receptors since the divergence of birds and mammals, as OR genes are found adjacent to *MINA* in eutherians but not in chicken. Olfactory receptors are the largest gene family in the mammalian genome and have evolved through intra- and inter-chromosomal duplications. At least some members of this family have been associated with evolutionary rearrangements (Yue & Haaf 2006). Duplicated regions, in general, are susceptible to illegitimate recombination, leading to large rearrangements. It is possible that the presence of OR genes adjacent to *MINA* has facilitated the recombination between physically distant loci, leading to the

relocation of *MINA* in wallaby to the subtelomeric region of 6q.

Likewise, *FTSJ1* maps to the X in eutherians in a region of highly conserved gene order. In other vertebrates, it maps near other human Xp genes. However, it is close to *HBB* on wallaby chromosome 5 and opossum chromosome 4 and is flanked by olfactory receptor genes at least in opossum. This implies that rearrangement occurred in a marsupial ancestor. A comparison of this same region between eutherians and opossum shows that the region has been completely rearranged (Figure 5). Three genes (*WDR13*, *GATA1* and *TBC1D25*) found on overlapping BACs on wallaby Xq reveal rearrangement of the region even between wallaby and opossum. This region of the eutherian X may have stabilized after divergence from marsupials owing to selection against rearrangements that would interrupt the spread of *XIST* (Mikkelsen *et al.* 2007). A search of the platypus genome assembly (Warren *et al.* 2008) shows that platypus *FTSJ1* is adjacent to other genes from the human X on platypus Ultra295 but all other genes in this region in human are on small contigs in platypus, making it difficult to know the gene order and chromosome location of genes within this region. *FTSJ1* and adjacent genes in human appear to be absent from the chicken genome assembly.

### *Evolution of human chromosomes*

Our comparison of wallaby chromosomes X and 5 with other vertebrate species has provided new insights into the evolution of the human X. One question we are now able to address is the location of the fusion point between the XAR and the XCR on the human X chromosome. Wilcox *et al.* (1996) proposed that the fusion point for these two regions was located within human Xp11.2 and that this fusion was followed by an inversion, as *SYN1* and *TBC1D25* (*OATL1*), which are flanked by XCR genes in human, were mapped by RISH to autosomes in the wallaby (Spencer *et al.* 1991a). Our mapping has assigned both of these genes to the long arm of the wallaby X, so it is unnecessary to propose an inversion event after the fusion of the XAR and XCR. We have narrowed down the fusion point on the human X to a 400 kb region between *RP2* (Xp11.4) within the XAR, and *RBM10* (Xq11.23) located in the XCR.

Wallaby chromosome 5 contains, in addition to genes from the human X, gene blocks from other human chromosomes, including one that contains genes that show imprinted expression. Comparisons of gene organization of these blocks between eutherians, marsupials and other vertebrates have provided insight into genome evolution as well as the origin of imprinting. In chicken, genes from the XAR, as well as from human chromosomes 2, 15 and 21, are part of one gene block on chicken chromosome 1 (Edwards *et al.* 2007). This block has been subject to several fission events to produce not only the XAR region, which maintains the same gene order as in chicken, but almost the entire human chromosome 21, a small part of human chromosome 2 and part of the Prader–Willi/Angelman syndrome imprinted region on chromosome 15 (Figure 6).

The human Prader–Willi/Angelman syndrome region consists of the paternally imprinted *UBE3A* gene associated with Angelman syndrome and the maternally silenced *SNRPN* gene linked to the Prader–Willi syndrome. Rapkins *et al.* (2006) showed that this imprinted region assembled after marsupial/eutherian divergence, as *UBE3A* is located on wallaby chromosome 5 and *SNRPN* on wallaby chromosome 1. Other imprinted genes from the region seem not to exist in marsupials. Several of these imprinted genes seem to have originated from RNA copies of genes at other locations in the genome. This represents the ancestral arrangement, since it is the same in chicken and platypus (Rapkins *et al.* 2006).

## Conclusions

We have devised an efficient strategy for mapping large conserved synteny blocks in non-model mammals, and applied this to generate dense maps of the X and 'neo-X' regions and to determine the arrangement of large conserved synteny blocks on chromosome 5. Comparisons between the wallaby and opossum chromosome maps revealed many rearrangements, highlighting the need for comparative gene mapping between South American and Australian marsupials. Of particular interest is the frequent rearrangement of the X which, along with the absence of a marsupial *XIST* gene, suggests that inactivation of the marsupial X chromosome may not depend on a whole-chromosome repression by a control locus.

## Acknowledgements

We thank Ke-Jun Wei for curation of the tammar wallaby BAC libraries, and Anthony Papenfuss and Kaighin McColl for bioinformatic tools used during this project. This project was funded by the Australian Research Council (to J.A.M.G. and M.B.R.).

## References

Alsop AE, Miethke P, Rofe R *et al.* (2005) Characterizing the chromosomes of the Australian model marsupial *Macropus eugenii* (tammar wallaby). *Chromosome Res* **13**: 627–636.

Bininda-Emonds OR, Cardillo M, Jones KE *et al.* (2007) The delayed rise of present-day mammals. *Nature* **446**: 507–512.

Brown CJ, Lafreniere RG, Powers VE *et al.* (1991) Localization of the X inactivation centre on the human X chromosome in Xq13. *Nature* **349**: 82–84.

Carrel L, Willard HF (2005) X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* **434**: 400–404.

Cooper DW, Johnston PG, Graves JAM (1993) X-inactivation in marsupials and monotremes. *Semin Dev Biol* **4**: 117–128.

Davidow LS, Breen M, Duke SE *et al.* (2007) The search for a marsupial XIC reveals a break with vertebrate synteny. *Chromosome Res* **15**: 137–146.

Deakin JE, Siddle HV, Cross JG, Belov K, Graves JAM (2007) Class I genes have split from the MHC in the tammar wallaby. *Cytogenet Genome Res* **116**: 205–211.

Delbridge ML, Graves JAM (2004) Assignment of the eukaryotic translation initiation factor (*EIF2S3*) to tammar wallaby chromosome 5p by in situ hybridization. *Cytogenet Genome Res* **107**: 139.

Delbridge ML, McMillan DA, Doherty RJ, Deakin JE, Graves JAM (2008) Origin and evolution of candidate mental retardation genes on the human X chromosome (MRX). *BMC Genomics* **9**: 65.

Duke SE, Samollow PB, Mauceli E, Lindblad-Toh K, Breen M (2007) Integrated cytogenetic BAC map of the genome of the gray, short-tailed opossum, *Monodelphis domestica*. *Chromosome Res* **15**: 361–370.

Duret L, Chureau C, Samain S, Weissenbach J, Avner P (2006) The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science* **312**: 1653–1655.

Edwards CA, Rens W, Clarke O *et al.* (2007) The evolution of imprinting: chromosomal mapping of orthologues of mammalian imprinted domains in monotreme and marsupial mammals—art. no. 157. *BMC Evol Biol* **7**: 157.

Fitzgerald J, Wilcox SA, Graves JAM, Dahl HH (1993) A eutherian X-linked gene, *PDHA1*, is autosomal in marsupials: a model for the evolution of a second, testis-specific variant in eutherian mammals. *Genomics* **18**: 636–642.

Foster JW, Graves JAM (1994) An *SRY*-related sequence on the marsupial X chromosome: implications for the evolution of the mammalian testis-determining gene. *Proc Natl Acad Sci U S A* **91**: 1927–1931.

Gartler SM, Dyer KA, Graves JAM, Rocchi M (1985) A two step model for mammalian X-chromosome inactivation. *Prog Clin Biol Res* **198**: 223–235.

Glas R, De Leo AA, Delbridge ML *et al.* (1999) Chromosome painting in marsupials: genome conservation in the kangaroo family. *Chromosome Res* **7**: 167–176.

Graves JAM (1995) The origin and function of the mammalian Y chromosome and Y-borne genes—an evolving understanding. *Bioessays* **17**: 311–320.

Graves JAM, Gartler SM (1986) Mammalian X chromosome inactivation: testing the hypothesis of transcriptional control. *Somat Cell Mol Genet* **12**: 275–280.

Graves JAM, Chew GK, Cooper DW, Johnston PG (1979) Marsupial–mouse cell hybrids containing fragments of the marsupial X chromosome. *Somatic Cell Genet* **5**: 481–489.

Heard E, Disteche CM (2006) Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes Dev* **20**: 1848–1867.

Hore TA, Koina E, Graves JAM (2007) The region homologous to the X-chromosome inactivation centre has been disrupted in marsupial and monotreme mammals. *Chromosome Res* **15**: 147–161.

Kirsch JAW, Lapointe FJ, Springer MS (1997) DNA-hybridisation studies of marsupials and their implications for metatherian classification. *Aust J Zool* **45**: 211–280.

Kohn M, Kehrer-Sawatzki H, Vogel W, Graves JAM, Hameister H (2004) Wide genome comparisons reveal the origins of the human X chromosome. *Trends Genet* **20**: 598–603.

Koina E, Graves JAM (2005) Assignment of the glucose-6-phosphate dehydrogenase (*G6PD*) gene to tammar wallaby chromosome Xq by fluorescence in situ hybridization with a BAC clone. *Cytogenet Genome Res* **108**: 362.

Koina E, Graves JAM (2006) Assignment of the proteolipid protein 1 gene (*PLP1*) to tammar wallaby chromosome Xq by fluorescence in situ hybridization with a BAC clone. *Cytogenet Genome Res* **114**: 94F.

Koina E, Wakefield MJ, Walcher C *et al.* (2005) Isolation, X location and activity of the marsupial homologue of *SLC16A2*, an XIST-flanking gene in eutherian mammals. *Chromosome Res* **13**: 687–698.

Mikkelsen TS, Wakefield MJ, Aken B *et al.* (2007) Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences. *Nature* **447**: 167–177.

National Human Genome Research Institute (2008) Internet references. Retrieved from http://www.genome.gov/12512299 1/9/2008.

Ohno S (1967) *Sex Chromosomes and Sex Linked Genes*. Berlin: Springer.

Pask A, Toder R, Wilcox SA, Camerino G, Graves JAM (1997) The candidate sex-reversing *DAX1* gene is autosomal in marsupials: implications for the evolution of sex determination in mammals. *Genomics* **41**: 422–426.

Pask A, Renfree MB, Graves JAM (2000) The human sex-reversing *ATRX* gene has a homologue on the marsupial Y chromosome, *ATRY*: implications for the evolution of mammalian sex determination. *Proc Natl Acad Sci U S A* **97**: 13198–13202.

Rapkins RW, Hore T, Smithwick M *et al.* (2006) Recent assembly of an imprinted domain from non-imprinted components. *PLoS Genet* **2**: e182.

Rens W, O'Brien PC, Fairclough H *et al.* (2003) Reversal and convergence in marsupial chromosome evolution. *Cytogenet Genome Res* **102**: 282–290.

Rofe R, Hayman D (1985) G-banding evidence for a conserved complement in the Marsupialia. *Cytogenet Cell Genet* **39**: 40–50.

Samollow PB (2006) Status and applications of genomic resources for the gray, short-tailed opossum, Monodelphis domestica, an American marsupial model for comparative biology. *Aust J Zool* **54**: 173–196.

Sharman GB (1971) Late DNA replication in the paternally derived X chromosome of female kangaroos. *Nature* **230**: 231–232.

Shetty S, Griffin DK, Graves JAM (1999) Comparative painting reveals strong chromosome homology over 80 million years of bird evolution. *Chromosome Res* **7**: 289–295.

Shevchenko AI, Zakharova IS, Elisaphenko EA *et al.* (2007) Genes flanking Xist in mouse and human are separated on the X chromosome in American marsupials. *Chromosome Res* **15**: 127–136.

Sinclair AH, Foster JW, Spencer JA *et al.* (1988) Sequences homologous to ZFY, a candidate human sex-determining gene, are autosomal in marsupials. *Nature* **336**: 780–783.

Spencer JA, Sinclair AH, Watson JM, Graves JAM (1991a) Genes on the short arm of the human X chromosome are not shared with the marsupial X. *Genomics* **11**: 339–345.

Spencer JA, Watson JM, Graves JAM (1991b) The X chromosome of marsupials shares a highly conserved region with eutherians. *Genomics* **9**: 598–604.

Spencer JA, Watson JM, Lubahn DB *et al.* (1991c) The androgen receptor gene is located on a highly conserved region of the X chromosomes of marsupial and monotreme as well as eutherian mammals. *J Hered* **82**: 134–139.

Svartman M, Vianna-Morgante AM (1998) Karyotype evolution of marsupials: from higher to lower diploid numbers. *Cytogenet Cell Genet* **82**: 263–266.

Toder R, Graves JAM (1998) *CSF2RA*, *ANT3*, and *STS* are autosomal in marsupials: implications for the origin of the pseudoautosomal region of mammalian sex chromosomes. *Mamm Genome* **9**: 373–376.

Tyndale-Biscoe H (2005) *Life of Marsupials*. Collingwood: CSIRO Publishing.

Tyndale-Biscoe CH, Renfree MB. 1987. *Reproductive Physiology of Marsupials*. Cambridge: Cambridge University Press.

Veyrunes F, Waters PD, Miethke P *et al.* (2008) Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. *Genome Res* **18**: 965–973.

Wakefield MJ, Graves JAM (2003) The kangaroo genome. Leaps and bounds in comparative genomics. *EMBO Rep* **4**: 143–147.

Wakefield MJ, Anderson M, Chang E *et al.* (2008) Cone visual pigments of monotremes: filling the phylogenetic gap. *Vis Neurosci* **25**: 257–264.

Warren WC, Hillier LW, Graves JAM *et al.* (2008) Genome analysis of the platypus reveals unique signatures of evolution. *Nature* **453**: 175–183.

Washington University Genome Sequencing Center (2008) Internet references. Retrieved from http://genome.wustl.edu/tools/software/overgo.cgi 1/9/2008.

Waters PD, Kirby PJ, Graves JAM (2001) Assignment of the *SMARCF1* gene to tammar wallaby chromosome 5q by fluorescence in-situ hybridization. *Cytogenet Cell Genet* **93**: 315–316.

Waters PD, Sankovic N, Kirby PJ, Delbridge ML, Graves JAM (2003) Assignment of the thymosin beta 4 X/Y chromosome (*TMSB4X/Y*) gene to tammar wallaby chromosome 5p by fluorescence in-situ hybridization. *Cytogenet Genome Res* **103**: 203F.

Watson JM, Spencer JA, Graves JAM, Snead ML, Lau EC (1992) Autosomal localization of the amelogenin gene in monotremes and marsupials: implications for mammalian sex chromosome evolution. *Genomics* **14**: 785–789.

Wilcox SA, Watson JM, Spencer JA, Graves JAM (1996) Comparative mapping identifies the fusion point of an ancient mammalian X-autosomal rearrangement. *Genomics* **35**: 66–70.

Woodburne MO, Rich TH, Springer MS (2003) The evolution of tribospheny and the antiquity of mammalian clades. *Mol Phylogenet Evol* **28**: 360–385.

Yu H, Pask AJ, Shaw G, Renfree MB (2006) Differential expression of *WNT4* in testicular and ovarian development in a marsupial. *BMC Dev Biol* **6**: 44.

Yue Y, Haaf T (2006) 7E olfactory receptor gene clusters and evolutionary chromosome rearrangements. *Cytogenet Genome Res* **112**: 6–10.

# APPENDIX 2: ORAL PRESENTATIONS

ANU Graduate Student Conference, Canberra, 8-9 October 2009

**Evolution of alpha and beta globin genes in amniotes**

Evolution- The Experience Conference,Melbourne, 8-13 February 2009

**Platypus globin genes; revisiting globin evolution**

The Institute for Genomics, Proteomics and Bioinformatics, Pennsylvania State University, USA, 27 August 2008

**A new model for the evolution of the globin gene cluster in amniotes**

Genetics Society of AustralAsia 55th Annual Conference, Adelaide, 7-10 July 2008

**A new model for the evolution of the globin gene cluster in amniotes**

Genetics Society of AustralAsia 54th Annual Conference, Sydney, 26-29 June 2007.

**Characterization of the beta-globin cluster and its flanking regions in the egg-laying monotreme, Ornithorhynchus anatinus (platypus)**,

KanGO Research Workshop, Melbourne, 21-22 August 2007

**Evolution of globin genes in mammals**

# APPENDIX 3: POSTER PRESENTATIONS

The 34th Lorne Genomics Conference, Mantra Erskine Resort, Lorne, 15-19 February 2009.

## Platypus globin genes; Revisiting globin evolution


Vidushi S. Patel[1], Stephen J.B. Cooper[2,3], Janine E. Deakin[1], Bob Fulton[4], Tina Graves[4], Wesley C. Warren[4], Richard K. Wilson[4], Jennifer A.M. Graves[1]

[1] The ARC Centre for Kangaroo Genomics, Research School of Biological Sciences, the Australian National University, Canberra Australia
[2] Australian Centre for Evolutionary Biology and Biodiversity, The University of Adelaide, Adelaide Australia
[3] Evolutionary Biology Unit, South Australian Museum, Adelaide Australia
[4] Genome Sequencing Centre, Washington University School of Medicine, St. Louis, United States of America


### Background

The alpha (α) and beta (β) globin gene clusters contain genes that together encode the polypeptide chains of haemoglobin. Although these genes have been extensively studied in vertebrates, questions about their evolution still remain. In amphibians, these two clusters are located together but in amniotes (birds and mammals), they lie on different chromosomes. The recent discovery of a fossil β-like globin gene (ω), beside the marsupial α-globin cluster, suggested that before the amniote radiation the entire α-β cluster duplicated onto another chromosome. This was followed by lineage-specific gene loss and duplication to produce paralogous α and β globin clusters in birds and mammals (Fig. 1). To test this hypothesis, and to explore α and β globin evolution at the stem of mammalian radiation, we have analyzed genomic data from an egg-laying monotreme mammal, the platypus (*Ornithorhynchus anatinus*). This information was used to characterize the platypus α and β globin clusters and their flanking regions, which then were compared with other vertebrate α and β globin clusters to gain insight into their evolution.

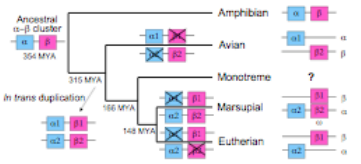



Fig.1 Previous model of globin evolution (Wheeler *et al.* 2001 & 2004)

### Results

#### 1. Characterisation of the platypus α and β globin clusters

The platypus α-globin cluster is located on chromosome 21 and contains eight genes; six α-like, one β-like (ω-globin), and a novel gene (Globin Y; *GBY*) (Fig. 2). The platypus β-globin cluster is located on chromosome 2 (2q5.1) and contains only two β-like globin genes (Fig. 2). Their identities were confirmed by phylogenetic analyses (Fig. 3).

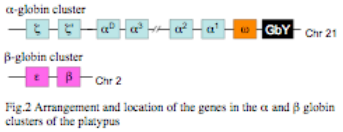



Fig.2 Arrangement and location of the genes in the α and β globin clusters of the platypus

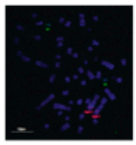


Fig.3 Phylogenetic trees of α-like and β-like globin genes from various vertebrates

#### 2. Analyses of flanking markers

In order to determine the genomic context of the α and β globin clusters, platypus orthologues of flanking genes were identified. The platypus α-globin cluster is flanked by *MPG*, *c16orf35* and *LUC7L*, whereas the β-globin cluster is embedded within many olfactory receptor genes (*ORG*). From the Ensembl database, the relative positions of these flanking genes along with *RRM1*, *CCKBR* and *ILK* to other vertebrate α and β globin clusters were analysed and summarized in Fig. 4.

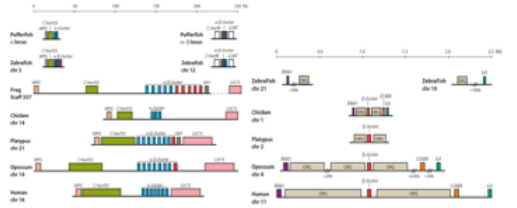


Fig.4 Arrangement of genes flanking the α and β globin clusters in different vertebrates

Locations of *MPG*, *c16orf35* and *LUC7L* are conserved adjacent to the α-globin cluster of amniotes, and in the same position adjacent to the α-β cluster of amphibians and, except *LUC7L*, in fish. However, the β-globin cluster of amniotes resides in a very different genomic context, sharing none of the flanking markers with the amniote α-globin cluster, or the α-β cluster of frog and fish. Instead, the amniote β-globin cluster is embedded in an array of *ORG* genes flanked by *RRM1*, *CCKBR* and *ILK* at conserved locations. These results and phylogenetic analyses (Fig. 3) refute the hypothesis that amniote α and β globin clusters evolved from different (paralogous) copies of an ancestral α-β region.

### Conclusion

Analysis of flanking markers implies that a very ancient region containing 5'-*MPG*-*C16orf35*-α-β-*GBY*-*LUC7L*-3' was present in the common ancestor of amniotes, frogs and fish, which has been conserved since the evolution of jawed vertebrates more than 450 million years ago (MYA). In amniotes, this ancient region has undergone some gene silencing to produce the α-globin cluster. The original ancestral β-globin gene (β1) within the α-globin cluster survives as the ω-globin gene in the monotreme and marsupial lineages. We propose that the amniote β-globin cluster (β2) evolved when a copy of an ancestral β-globin gene was transposed into a different chromosome within a region replete with multiple copies of *ORG* genes, before the divergence of amniotes 310 MYA (Fig. 5). This transposed β-globin gene (β2) duplicated several times to form different clusters in different amniote lineages (Patel *et al.* 2008).


**References:**

Wheeler et al: **An orphaned mammalian beta-globin gene of ancient evolutionary origin**. *Proc Natl Acad Sci USA* 2001, **98**: 1101-1106

Wheeler et al: **Linkage of the beta-like omega-globin gene to alpha-like globin genes in an Australian marsupial supports the chromosome duplication model for separation of globin gene clusters**. *J Mol Evol* 2004, **58**: 642-652

Patel et al: **Platypus globin genes and flanking loci suggest a new insertional model for beta-globin evolution in birds and mammals**. *BMC Biology* 2008, **6**: 34

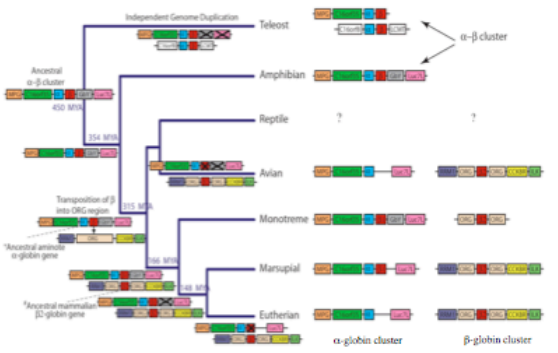




Fig.5 Our new model of globin evolution in vertebrates

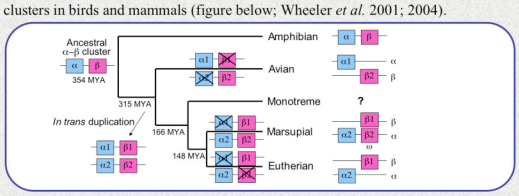
# Platypus globin genes and flanking loci suggest a new insertional model for β-globin evolution in birds and mammals

Vidushi S. Patel[1], Stephen J.B. Cooper[2,3], Janine E. Deakin[1], Bob Fulton[4], Tina Graves[4], Wesley C. Warren[4], Richard K. Wilson[4], Jennifer A.M. Graves[1]

[1] *The ARC Centre for Kangaroo Genomics, Research School of Biological Sciences, the Australian National University, Canberra Australia.* [2] *Australian Centre for Evolutionary Biology and Biodiversity, The University of Adelaide, Adelaide Australia.* [3] *Evolutionary Biology Unit, South Australian Museum, Adelaide Australia.* [4] *Genome Sequencing Centre, Washington University School of Medicine, St. Louis, United States*

## Background

The alpha (α) and beta (β) globin gene clusters contain genes that together encode the polypeptide chains of haemoglobin. Although these genes have been extensively studied in vertebrates, questions about their evolution still remain. In amphibians, these two clusters are located together but in amniotes (birds and mammals), they lie on different chromosomes. The recent discovery of a fossil β-like globin gene (ω), beside the marsupial α-globin cluster, suggested that before the amniote radiation the entire α-β cluster duplicated onto another chromosome. This was followed by lineage-specific gene loss and duplication to produce paralogous α and β globin clusters in birds and mammals (figure below; Wheeler *et al.* 2001; 2004).




To test this hypothesis, and to explore α and β globin evolution at the stem of mammalian radiation, we have analyzed genomic data from an egg-laying monotreme mammal, the platypus (*Ornithorhynchus anatinus*).

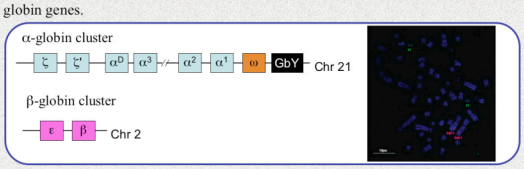
## Aims

- to characterise the platypus α and β globin clusters
- to analyze the flanking region of vertebrate α and β globin clusters
- to study α and β globin evolution in vertebrates

## Results

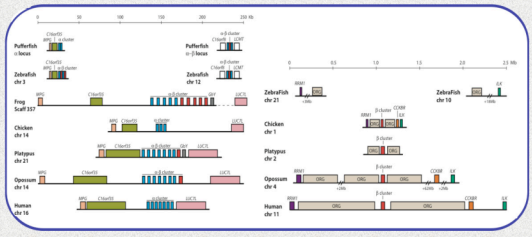
### 1. Characterisation of the platypus α and β globin clusters

The platypus α-globin cluster is located on chromosome 21 and contains eight genes; six α-like, one β-like (ω-globin), and a novel gene (Globin Y; *GBY*). The platypus β-globin cluster is located on chromosome 2 (2q5.1) and contains only two β-like globin genes.

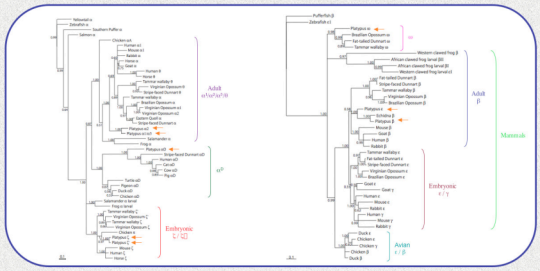



### 2. Analyses of flanking markers

In order to determine the genomic context of the α and β globin clusters, platypus orthologues of flanking genes were identified. The platypus α-globin cluster is flanked by *MPG*, *c16orf35* and *LUC7L*, whereas the β-globin cluster is embedded within many olfactory receptor genes (*ORG*). From the Ensembl database, the relative positions of these flanking genes along with *RRM1*, *CCKBR* and *ILK* to other vertebrate α and β globin clusters were analysed and summarised.



Locations of *MPG*, *c16orf35* and *LUC7L* are conserved adjacent to the α-globin cluster of amniotes, and in the same position adjacent to the α-β cluster of amphibians and, except *LUC7L*, in fish. However, the β-globin cluster of amniotes resides in a very different genome context, sharing none of the flanking markers with the amniote α-globin cluster, or the α-β cluster of frog and fish. Instead, the amniote β-globin cluster is embedded in an array of *ORG* genes flanked by *RRM1*, *CCKBR* and *ILK* at conserved locations. These results and phylogenetic analyses (shown below) refutes the hypothesis that amniote α and β globin clusters evolved from different (paralogous) copies of an ancestral α-β region.

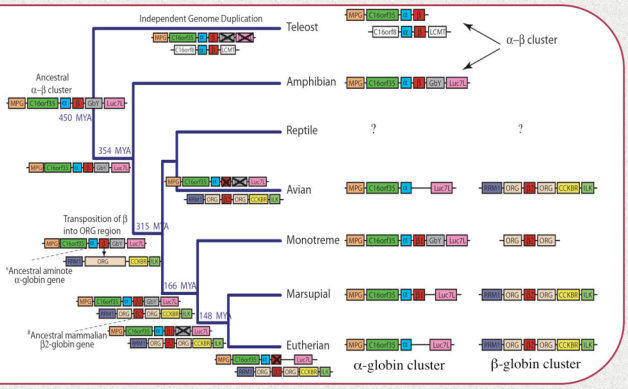


## Conclusion

### Our new model of globin evolution in vertebrates

Analysis of flanking markers implies that a very ancient region containing 5'-*MPG-C16orf35*-α-β-*GBY-LUC7L*-3' was present in the common ancestor of amniotes, frogs and fish, which has been conserved since the evolution of jawed vertebrates more than 450 million years ago (MYA). In amniotes, this ancient region has undergone some gene silencing to produce the α-globin cluster. The original ancestral β-globin gene (β1) within the α-globin cluster survives as the ω-globin gene in the monoteme and marsupial lineages. We propose that the amniote β-globin cluster (β2) evolved when a copy of an ancestral β-globin gene was transposed into a different chromosome within a region replete with multiple copies of *ORG* genes, before the divergence of amniotes 310 MYA. This transposed β-globin gene (β2) duplicated several times to form different clusters in different amniote lineages.

(Patel *et al.* 2008. BMC Biology)




References: Wheeler, *et al.* 2001. *Proc Natl Acad Sci USA* **98**: 1101-1106; Wheeler, *et al.* 2004. *J Mol Evol* **58**: 642-652

The 33rd Lorne Genomics Conference, Mantra Erskine Resort, Lorne, 16-21 February 2008.

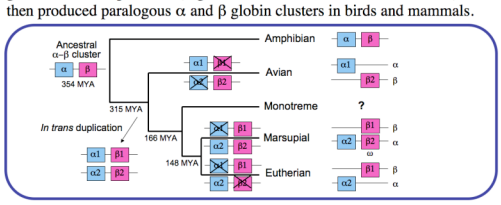
# Characterisation of the haemoglobin clusters and their flanking regions in the egg-laying monotreme, *Ornithorhynchus anatinus* (platypus)


V. S. Patel[1*], S. J.B. Cooper[2,3], J. E. Deakin[1], B. Fulton[4], T. Graves[4], W. C. Warren[4], R. K. Wilson[4], J. A.M. Graves[1]

[1] *Research School of Biological Sciences, the Australian National University, Canberra Australia*
[2] *Australian Centre for Evolutionary Biology and Biodiversity, The University of Adelaide, Adelaide Australia*
[3] *Evolutionary Biology Unit, South Australian Museum, Adelaide Australia*
[4] *Genome Sequencing Centre, Washington University School of Medicine, St. Louis, United States*


## Background

The alpha (α) and beta (β) globin gene clusters contain genes that together encode the polypeptide chains of haemoglobin. Although these genes have been extensively studied in vertebrates, questions about their evolution still remain. In amphibians, these two clusters are located together but in amniotes (birds and mammals), they appear on different chromosomes. The recent discovery of a fossil β-like globin gene (ω) beside the marsupial α-globin cluster, suggested that before the amniote radiation the entire α-β cluster duplicated onto another chromosome, followed by lineage-specific gene loss and duplication (Figure below; Wheeler *et al*. 2001; 2004). This then produced paralogous α and β globin clusters in birds and mammals.




To test this hypothesis and to explore haemoglobin evolution at the stem of mammalian radiation, we have analyzed genomic data from an egg-laying monotreme mammal, the platypus (*Ornithorhynchus anatinus*).

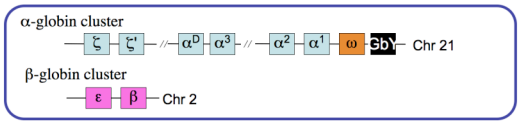
## Aims

1. Characterise the platypus α and β globin clusters
2. Analyze genes that flank the vertebrate α and β globin clusters
3. Study haemoglobin evolution in vertebrates

## Results

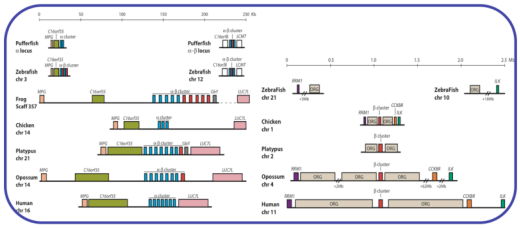
### Characterisation of the platypus α- and β- globin clusters

The platypus α-globin cluster is located on chromosome 21 and contains a total of eight genes; six α-like, one β-like (ω-globin), and a novel gene (Globin Y; GbY). The platypus β-globin cluster is located on chromosome 2 (2q5.1) and contains only two β-like globin genes.




### Analyses of flanking markers

In order to determine the genomic context of the α- and β- globin clusters, flanking genes were identified. The platypus α-globin cluster is flanked by *MPG*, *c16orf35* and *Luc7L*, while the β-globin cluster is flanked by many olfactory receptor genes (*ORG*). From the Ensembl database, the relative positions of these flanking genes along with *RRM1*, *CCKBR* and *ILK* to other vertebrate α- and β- globin clusters were analysed and summarised.

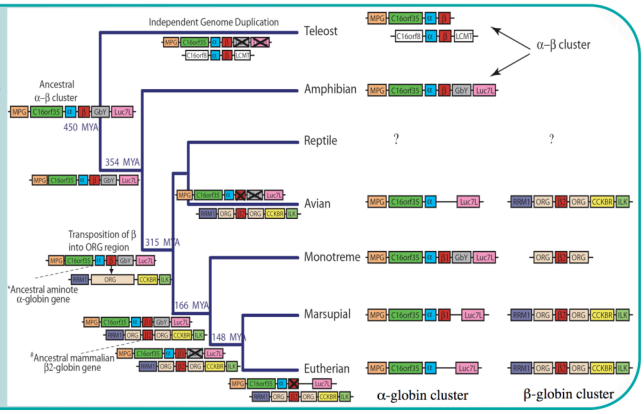


Locations of *MPG*, *c16orf35* and *Luc7L* are conserved adjacent to the α-globin cluster of amniotes and in the same position adjacent to the α-β cluster of amphibians and, except *Luc7L*, in fish. However, the β-globin cluster of amniotes resides in a very different genome context, sharing none of the flanking markers with the amniote α-globin cluster, or the α-β cluster of frog and fish. Instead, the amniote β-globin cluster is embedded in an array of olfactory receptor genes (*ORG*) and flanked by *RRM1*, *CCKBR* and *ILK* at conserved locations. These results and phylogenetic analyses (data not shown) refutes the hypothesis that amniote α- and β- globin clusters evolved from different (paralogous) copies of an ancestral α-β region.

## Conclusion

### Our Model of Haemoglobin Evolution in Vertebrates

The flanking analyses suggest that a very ancient region containing 5'-*MPG-C16orf35*-α-β-*GbY-Luc7L*-3' was present in the common ancestor of amniote, frog and fish, which has been conserved since the evolution of jawed vertebrates more than 450 MYA. In amniotes, this ancient region has undergone some gene silencing and represents the so called α-globin cluster. The original ancestral β-globin gene (β1) within the α-globin cluster is represented by the platypus and marsupial ω-globin gene, which may have been lost in aves and eutherians. We hypothesize that the amniote β-globin cluster (β2) has evolved by a copy of an ancestral β-globin gene being inserted into a different chromosome within a region replete with multiple copies of *ORG* genes before the divergence of amniotes 310 MYA. This transposed β-globin gene (β2) duplicated several times to form different clusters in the different lineages.





References:
Wheeler, *et al*. 2001. *Proc Natl Acad Sci USA* **98**: 1101-1106
Wheeler, *et al*. 2004. *J Mol Evol* **58**: 642-652



*Corresponding Author: vidushi.patel@anu.edu.au

Patel *et al*. 2008. **BMC Genomics**. Submitted